

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 13:59:39 ; Search time 492 Seconds
(without alignments)
436.049 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 36

Sequence: 1 EPHHHGQGMHGYDADVRLYRRHHGSGSPSRHRR 36

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5532208 seqs, 2979665951 residues

Word size: 1

Total number of hits satisfying chosen parameters: 11088970

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=olp
-MODEL=frame_p2n.model -DEV=olp
-Q/cgn2_1/USPTO.spool_p/US10500018/runat_25032005_092651_24611/app_query.fasta_1.199
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40 -LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10500018 -CGN 1 1 723 @runat_25032005_092651_24611
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FPGAPOP=6 -FPGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
- 14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:
- 16: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:
- 17: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq:
- 18: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:
- 19: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
- 20: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:
- 21: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
- 22: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	8	22.2	292	8	US-08-781-986A-4465	Sequence 4465, Ap
c 2	8	22.2	292	17	US-10-329-624-4465	Sequence 4465, Ap
c 3	8	22.2	424	18	US-10-425-115-177211	Sequence 177211, A
c 4	8	22.2	522	18	US-10-437-963-39610	Sequence 39610, A
c 5	8	22.2	725	13	US-10-027-632-144873	Sequence 144873, A
c 6	8	22.2	725	13	US-10-027-632-144873	Sequence 144873, A
c 7	8	22.2	857	18	US-10-425-115-146096	Sequence 146096, A
c 8	8	22.2	939	15	US-10-259-165-709	Sequence 709, App
c 9	8	22.2	939	17	US-10-260-238-5471	Sequence 5471, App
c 10	8	22.2	1274	18	US-10-767-701-10615	Sequence 10615, A
c 11	8	22.2	1279	18	US-10-437-963-69415	Sequence 69415, A
c 12	8	22.2	1953	17	US-10-282-122A-7912	Sequence 7912, Ap
c 13	8	22.2	2031	18	US-10-437-963-14823	Sequence 14823, A
c 14	8	22.2	2277	18	US-10-437-963-61184	Sequence 61184, A
c 15	8	22.2	2465	18	US-10-425-115-152186	Sequence 152186, A
c 16	8	22.2	3213	17	US-10-369-493-42962	Sequence 42962, A
c 17	8	22.2	5002	18	US-10-437-963-53269	Sequence 53269, A
c 18	8	22.2	9623	8	US-08-781-986A-166	Sequence 166, App
c 19	8	22.2	9623	17	US-10-329-624-166	Sequence 166, App
c 20	8	22.2	17493	19	US-10-211-028-10	Sequence 10, Appl
c 21	8	22.2	90597	19	US-10-211-028-10	Sequence 10, Appl
c 22	7	19.4	25	19	US-10-719-900-351970	Sequence 351970, A
c 23	7	19.4	40	18	US-10-469-851-137	Sequence 137, App
c 24	7	19.4	45	16	US-10-057-531A-11	Sequence 11, Appl
c 25	7	19.4	45	16	US-10-057-532A-11	Sequence 11, Appl
c 26	7	19.4	57	16	US-10-057-531A-12	Sequence 12, Appl
c 27	7	19.4	57	16	US-10-057-532A-12	Sequence 12, Appl
c 28	7	19.4	100	15	US-10-057-810-33	Sequence 33, Appl
c 29	7	19.4	100	15	US-10-057-828-33	Sequence 33, Appl
c 30	7	19.4	124	9	US-09-815-343-818	Sequence 818, App
c 31	7	19.4	124	9	US-09-922-217-1024	Sequence 1024, App
c 32	7	19.4	124	9	US-09-833-263-1024	Sequence 1024, App
c 33	7	19.4	124	13	US-10-025-380-1024	Sequence 1024, App
c 34	7	19.4	124	17	US-10-097-105-818	Sequence 818, App
c 35	7	19.4	137	18	US-10-425-115-61631	Sequence 61631, A
c 36	7	19.4	186	17	US-10-242-535A-2395	Sequence 2395, Ap
c 37	7	19.4	186	17	US-10-085-783A-2395	Sequence 2395, Ap
c 38	7	19.4	198	15	US-10-156-761-1079	Sequence 1079, Ap
c 39	7	19.4	201	18	US-10-741-601-21795	Sequence 21795, A
c 40	7	19.4	201	18	US-10-719-993-49164	Sequence 49164, A
c 41	7	19.4	201	18	US-10-719-993-49173	Sequence 49173, A
c 42	7	19.4	201	18	US-10-719-993-50002	Sequence 50002, A
c 43	7	19.4	201	18	US-10-719-993-50005	Sequence 50005, A
c 44	7	19.4	201	19	US-10-741-600-61102	Sequence 61102, A
c 45	7	19.4	201	19	US-10-741-600-61103	Sequence 61103, A

ALIGNMENTS

RESULT 1

US-08-781-986A-4465/c
; Sequence 4465, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
; Sequence 4465, Diskette, 3.50 inch, 1.4Mb storage
; MEDIUM TYPE: HP Vectra 486/33
; COMPUTER: MSDOS version 6.2
; OPERATING SYSTEM: ASCII Text
; SOFTWARE: ASCII Text

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4465:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-4465

Alignment Scores:
Pred. No.: 44 Length: 292
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.22% Indels: 0
DB: 8 Gaps: 0

US-10-500-018A-27 (1-36) x US-08-781-986A-4465 (1-292)
QY 19 ValatgLeuTyArGaArqHis 26
Db 237 GTGGGACTCTATGCCGACCCAC 214

RESULT 2
US-10-329-624-4465/c
; Sequence 4465, Application US/10329624
; Publication No. US20040043037A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman

;
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 4465:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-10-329-624-4465

Alignment Scores:
Pred. No.: 44 Length: 292
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.22% Indels: 0
DB: 17 Gaps: 0

US-10-500-018A-27 (1-36) x US-10-329-624-4465 (1-292)
QY 19 ValatgLeuTyArGaArqHis 26
Db 237 GTGGGACTCTATGCCGACCCAC 214

RESULT 3
US-10-425-115-177211
; Sequence 177211, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 177211
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (424)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_93205C.1
; US-10-425-115-177211

Alignment Scores:
Pred. No.: 57.9 Length: 424
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.22% Indels: 0
DB: 18 Gaps: 0

US-10-500-018A-27 (1-36) x US-10-425-115-177211 (1-424)
QY 3 HighHisHisProGlyArg 10
Db 262 CACCACACCCACCCCGAGGCCGA 285

RESULT 4
US-10-437-963-39610
; Sequence 39610, Application US/10437963
; Publication No. US20040123343A1
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 13:24:53 ; Search time 133 Seconds
(without alignments)
442.902 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 36

Sequence: 1 EPHHHPGGQMGYDADVRLRRHHGSGSPSRHR 36

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2397488

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10500018/runat_25032005_092650_24528/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=olip2n.rnl -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10500018 @CGN 1.1.105 @runat_25032005_092650_24528 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT_DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTC.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22.2	292	4	US-08-956-171E-4465	Sequence 4465, Ap
C 2	22.2	292	4	US-08-781-986A-4465	Sequence 4465, Ap
C 3	22.2	594	4	US-09-252-991A-15947	Sequence 15947, A
C 4	22.2	720	4	US-09-252-991A-8676	Sequence 8676, Ap
C 5	22.2	921	4	US-09-252-991A-8624	Sequence 8624, Ap
C 6	22.2	1215	4	US-09-252-991A-16443	Sequence 16443, A
C 7	22.2	1524	4	US-09-252-991A-16154	Sequence 16154, A
C 8	22.2	1656	4	US-09-902-540-6836	Sequence 6836, Ap
C 9	22.2	2616	4	US-09-252-991A-16042	Sequence 16042, A
C 10	22.2	2739	4	US-09-252-991A-16558	Sequence 16558, A
C 11	22.2	3252	4	US-09-902-540-9254	Sequence 9254, Ap
C 12	22.2	4026	4	US-09-902-540-571	Sequence 571, App

13	8	22.2	9623	4	US-08-956-171E-166	Sequence 166, App
14	8	22.2	9623	4	US-08-781-986A-166	Sequence 166, App
15	8	22.2	10322	4	US-09-902-540-989	Sequence 989, App
C 16	7	19.4	30	2	US-08-489-269-1	Sequence 1, Appl1
C 17	7	19.4	39	1	US-08-352-179-8	Sequence 8, Appl1
C 18	7	19.4	39	1	US-08-352-179-9	Sequence 9, Appl1
C 19	7	19.4	99	1	US-08-352-179-1	Sequence 1, Appl1
C 20	7	19.4	191	1	US-07-616-022C-4	Sequence 4, Appl1
C 21	7	19.4	237	2	US-08-530-290-6	Sequence 6, Appl1
C 22	7	19.4	274	3	US-09-275-848-9	Sequence 9, Appl1
C 23	7	19.4	274	3	US-09-275-848-10	Sequence 10, Appl1
C 24	7	19.4	274	4	US-09-867-345-9	Sequence 9, Appl1
C 25	7	19.4	274	4	US-09-867-345-10	Sequence 10, Appl1
C 26	7	19.4	287	4	US-09-313-294A-4178	Sequence 4178, Ap
C 27	7	19.4	322	3	US-09-098-287A-6	Sequence 6, Appl1
C 28	7	19.4	322	4	US-09-445-649-6	Sequence 6, Appl1
C 29	7	19.4	331	4	US-09-468-253B-17	Sequence 17, Appl1
C 30	7	19.4	420	4	US-09-468-253B-20	Sequence 20, Appl1
C 31	7	19.4	451	1	US-08-142-551B-128	Sequence 128, App
C 32	7	19.4	558	4	US-09-252-991A-4406	Sequence 4406, App
C 33	7	19.4	591	4	US-09-252-991A-13345	Sequence 13345, A
C 34	7	19.4	601	4	US-09-949-016-22453	Sequence 22453, A
C 35	7	19.4	601	4	US-09-949-016-22454	Sequence 22454, A
C 36	7	19.4	601	4	US-09-949-016-22455	Sequence 22455, A
C 37	7	19.4	601	4	US-09-949-016-150101	Sequence 150101, A
C 38	7	19.4	601	4	US-09-949-016-150102	Sequence 150102, A
C 39	7	19.4	601	4	US-09-949-016-154258	Sequence 154258, A
C 40	7	19.4	601	4	US-09-949-016-154259	Sequence 154259, A
C 41	7	19.4	601	4	US-09-949-016-154260	Sequence 154260, A
C 42	7	19.4	601	4	US-09-949-016-161888	Sequence 161888, A
C 43	7	19.4	675	4	US-09-902-540-3775	Sequence 3775, Ap
C 44	7	19.4	710	3	US-09-098-287A-5	Sequence 5, Appl1
C 45	7	19.4	710	4	US-09-445-649-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1

US-08-956-171E-4465/c

; Sequence 4465, Application US/08956171E

; Patent No. 6593114

; GENERAL INFORMATION:

APPLICANT: Charles Kunsch

Gil H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439
 INFORMATION FOR SEQ ID NO: 4465:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 292 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 4465:

US-08-956-171E-4465
 Alignment Scores:
 Pred. No.: 20 Length: 292
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 22.22% Indels: 0
 DB: 4 Gaps: 0

US-10-500-018A-27 (1-36) x US-08-956-171E-4465 (1-292)

Qy 19 ValArgLeuTyArgArgHis 26
 Db 237 GTGCGACTCTATCGCCGACACCAC 214

RESULT 2

US-08-781-986A-4465/c
 Sequence 4465, Application US/08781986A
 Patent No. 6737248
 GENERAL INFORMATION:

APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,986A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob
 REGISTRATION NUMBER: 30,446
 REFERENCE/DOCKET NUMBER: PB248PP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 4465:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 292 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-781-986A-4465

Alignment Scores:
 Pred. No.: 20 Length: 292
 Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 22.22% Indels: 0
 DB: 4 Gaps: 0

US-10-500-018A-27 (1-36) x US-08-781-986A-4465 (1-292)

Qy 19 ValArgLeuTyArgArgHis 26
 Db 237 GTGCGACTCTATCGCCGACACCAC 214

RESULT 3

US-09-252-991A-15947/c
 Sequence 15947, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 15947
 LENGTH: 594
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-15947

Alignment Scores:
 Pred. No.: 37.7 Length: 594
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 22.22% Indels: 0
 DB: 4 Gaps: 0

US-10-500-018A-27 (1-36) x US-09-252-991A-15947 (1-594)

Qy 23 ArgArgHisGlySerGlySer 30
 Db 560 CGGCGACACACCGGAGTGTGTAGT 537

RESULT 4

US-09-252-991A-8676/c
 Sequence 8676, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 8676
 LENGTH: 720
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-8676

Alignment Scores:
 Pred. No.: 44.8 Length: 720
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 13:17:03 ; Search time 3086 Seconds
(without alignments)
444.042 Million cell updates/sec

Title: US-10-500-018a-27
Perfect score: 36
Sequence: 1 EPHHHPGQRMHGYDADVRLYRRHGGSPSRHRR 36

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68473561

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10500018/runat_25032005_092650_24505/app_query.fasta_1.199
-DB=EST -Qfmt=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us10500018 @CGN 1.1 5180 @runat_25032005_092650_24505 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	25.0	215	6	CA352397
2	9	25.0	215	6	CA365706
3	9	25.0	288	2	BF466257
4	9	25.0	328	1	AI852035
5	9	25.0	429	9	CE113447
6	9	25.0	577	5	BF63227
7	9	25.0	610	1	AI587738
8	9	25.0	635	6	C76297
9	9	25.0	645	5	BU239570

C	10	9	25.0	681	5	BU121670
C	11	9	25.0	725	5	BP762221
C	12	9	25.0	857	2	BF676868
C	13	9	25.0	864	9	CR272725
C	14	9	25.0	963	7	CF577704
C	15	9	25.0	1428	5	BU105990
C	16	9	25.0	1671	3	BC022694
C	17	8	22.2	227	2	AW880150
C	18	8	22.2	271	2	BE409392
C	19	8	22.2	277	7	CF860372
C	20	8	22.2	303	2	AW122059
C	21	8	22.2	303	2	BE950947
C	22	8	22.2	310	2	BE956312
C	23	8	22.2	374	6	CA002052
C	24	8	22.2	387	2	BF286330
C	25	8	22.2	419	2	BF293549
C	26	8	22.2	428	5	BQ467112
C	27	8	22.2	433	2	AW336993
C	28	8	22.2	468	7	CF757810
C	29	8	22.2	480	6	CB209956
C	30	8	22.2	491	4	BI720481
C	31	8	22.2	491	6	CA594554
C	32	8	22.2	497	4	BI997490
C	33	8	22.2	501	4	BJ196708
C	34	8	22.2	504	7	CF575893
C	35	8	22.2	510	8	AQ321319
C	36	8	22.2	513	5	BP099995
C	37	8	22.2	517	4	BM098004
C	38	8	22.2	521	4	BG300210
C	39	8	22.2	524	7	CF761905
C	40	8	22.2	527	9	CE110903
C	41	8	22.2	528	2	AW294076
C	42	8	22.2	531	8	BH620889
C	43	8	22.2	531	8	BH622459
C	44	8	22.2	535	6	CA237515
C	45	8	22.2	537	1	AV432534

ALIGNMENTS

RESULT 1
CA352397
LOCUS

CA352397 215 bp mRNA linear EST 05-NOV-2002
623644 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT60010_C_H05 5', mRNA sequence.

CA352397 GI:24597568
CA352397.1 Oncorhynchus mykiss (rainbow trout) EST.
SOURCE
ORGANISM

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus. 1 (bases 1 to 215)

REXROAD, C.E. 3rd, Lee, Y., Keele, J.W., Karamycheva, S., Brown, G., Koop, B., Gahr, S.A., Palti, Y. and Quackenbush, J.
Sequence analysis of a rainbow trout cDNA library and creation of a gene index
Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
Contact: REXROAD CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351

FEATURES

source
1..215
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
Email: crexroad@cccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified by cross match v0.990329.
Seq primer: ACGGATACAAATTCACACAGA.
Location/Qualifiers

/db_xref="taxon:8022"
 /clone="IRT60010_C_H05"
 /tissue_type="pooled"
 /lab_host="DH108"
 /clone_lib="NCCCWA lrt"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."

ORIGIN

Alignment Scores:
 Pred. No.: 35.6 Length: 215
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.00% Indels: 0
 DB: 6 Gaps: 0

US-10-500-018A-27 (1-36) x CA352397 (1-215)

Qy 27 GlySerGlySerProSerArgHisArg 35
 |||||
 Db 164 GGCTCTGGAAGTCTTCTCGTCATAGA 190

RESULT 2

CA365706 640899 NCCCWA lrt Oncorhynchus mykiss cdna linear EST 06-NOV-2002
 LOCUS 5', mRNA
 DEFINITION CA365706.1 GI:24675924

ACCESSION CA365706
 VERSION 1
 KEYWORDS EST
 SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE

AUTHORS 1 (bases 1 to 215)
 TITLE Sequence analysis of a rainbow trout cdna library and creation of a
 JOURNAL gene index
 COMMENT Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
 Contact: Rexroad CE
 USDA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Leetown Road, Kearneysville, WV 25430, USA
 Tel: 304 724 8340 x2129
 Fax: 304 725 0351
 Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified by
 cross match v0.990329.
 Seq primer: AGCGGATACACATTTTCACACAGGA.

FEATURES

source
 1..215
 Location/Qualifiers
 /organism="Oncorhynchus mykiss"
 /mol_type="mRNA"
 /db_xref="taxon:8022"
 /clone="IRT115L21_B_F11"
 /tissue_type="pooled"
 /lab_host="DH108"
 /clone_lib="NCCCWA lrt"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."

ORIGIN

Alignment Scores:
 Pred. No.: 35.6 Length: 215
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.00% Indels: 0

DB: 6 Gaps: 0

US-10-500-018A-27 (1-36) x CA365706 (1-215)

Qy 27 GlySerGlySerProSerArgHisArg 35
 |||||
 Db 164 GGCTCTGGAAGTCTTCTCGTCATAGA 190

RESULT 3

BF466257 288 bp mRNA linear EST 04-DEC-2000
 LOCUS UI-M-CGOp-bqv-c-12-0-UI.s1 NIH BMAP Ret4 S2 Mus musculus cdna clone
 DEFINITION UI-M-CGOp-bqv-c-12-0-UI 3', mRNA sequence.

ACCESSION BF466257
 VERSION 1
 KEYWORDS EST
 SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 288)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 COMMENT Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

REFERENCE

AUTHORS Contact: Chin, H
 TITLE National Institute of Mental Health
 JOURNAL 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 MEDLINE 20892-9643, USA
 PUBMED Tel: 301 443 1706
 COMMENT Fax: 301 443 9890
 Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 embryonic retina tissue cDNA Library Preparation. M.B. Soares Lab
 Clone distribution: Researchers may obtain BMAP cDNA clones from
 RESEARCH GENETICS. It should be noted that Bento Soares is
 generating a small number of additional specialized non-redundant
 arrays of BMAP cDNAs whose availability will be considered under
 appropriate and limited collaborative arrangements. The following
 restrictive elements were found in this cDNA sequence: 1-117,
 >PB1D#SINE/Alu
 Seq primer: M13 Forward
 POUfA-res.

FEATURES

source
 1..288
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-CGOp-bqv-c-12-0-UI"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NIH BMAP Ret4 S2"
 /note="Vector: pT7T3D-pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The
 NIH BMAP Ret4 S2 library is a subtracted library,
 ultimately derived from mouse retina tissue libraries at
 various stages of development. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu. The tissue
 for this library was contributed by Dr. Xin-Yuan Fu, Yale
 University School of Medicine
 TAG TISSUE=embryonic-retina
 TAG LIB=NIH BMAP Ret4_S2
 TAG_SEQ=CTGCTAGTAGCA"

ORIGIN

Alignment Scores:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 11:58:13 ; Search time 430 Seconds
(without alignments)
495.606 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 36

Sequence: 1 EPHHHPGGRQMGGYDADVRLYRRHGGSPSRHRR 36

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8770668

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q=/cfn2_1/USPTO_spool_p/US10500018/runat_25032005_092649_24483/app_query.fasta_1.199
-DB=N_Geneseq_16Dec04 -QFWT=fatasp -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10500018 @CGN 1.1.708 @runat_25032005_092649_24483 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	41.7	627	2 AAX99247	Aax99247 Nucleotid
2	10	27.8	36	12 ADL66849	Adl66849 Ghrelin a
C 3	10	27.8	55	2 AAX99259	Aax99259 Nucleotid
C 4	10	27.8	55	3 AAA99157	Aaa99157 Ompt prot
5	9	25.0	47	2 AAX99242	Aax99242 Nucleotid

6	9	25.0	47	3	AAA99156	Aaa99156 Ompt prot
7	8	22.2	30	3	AAA99174	Aaa99174 Ompt prot
C 8	22.2	292	2	2	AAV78776	Aav78776 Staphyloc
9	8	22.2	474	4	AAI88297	Aai88297 Human pol
C 10	8	22.2	594	11	ABD17343	Abd17343 Pseudomon
C 11	8	22.2	720	11	ABD10072	Abd10072 Pseudomon
C 12	8	22.2	921	11	ABD10020	Abd10020 Pseudomon
C 13	8	22.2	939	11	ADJ12073	Adj12073 Maizo CDN
14	8	22.2	939	12	ADJ44471	Adj44471 Plant CDN
15	8	22.2	1215	11	ABD17839	Abd17839 Pseudomon
C 16	8	22.2	1524	11	ABD17550	Abd17550 Pseudomon
C 17	8	22.2	1950	8	ACF74273	Acf74273 Staphyloc
C 18	8	22.2	1953	8	ACA20042	Ac20042 Prokaryot
C 19	8	22.2	2111	2	AAT80389	Aat80389 Staphyloc
C 20	8	22.2	2616	11	ABD17438	Abd17438 Pseudomon
C 21	8	22.2	2739	11	ABD17954	Abd17954 Pseudomon
C 22	8	22.2	3213	13	ADT44524	Adt44524 Bacterial
C 23	8	22.2	3282	4	AAS59652	Aas59652 Propionib
C 24	8	22.2	3282	8	ACF64581	Acf64581 Propionib
C 25	8	22.2	9623	2	AAV74477	Aav74477 Staphyloc
26	8	22.2	90597	10	ADJ72363	Adj72363 Streptomy
C 27	8	22.2	90600	6	ABQ78872	Abq78872 S. roseos
C 28	8	22.2	154681	10	ADL13861	Adl13861 Osteoarth
C 29	7	19.4	30	3	AAA99173	Aaa99173 Ompt prot
C 30	7	19.4	30	3	AAA99162	Aaa99162 Ompt prot
C 31	7	19.4	30	3	AAA99172	Aaa99172 Ompt prot
C 32	7	19.4	30	3	AAA99176	Aaa99176 Ompt prot
C 33	7	19.4	30	3	AAA99161	Aaa99161 Ompt prot
C 34	7	19.4	30	3	AAA99175	Aaa99175 Ompt prot
C 35	7	19.4	39	2	AAQ36550	Aaq36550 Encodes A
C 36	7	19.4	40	6	ABT12116	Abt12116 E coli ex
C 37	7	19.4	41	2	AAV14056	Aav14056 Primer us
C 38	7	19.4	45	6	ABK89063	Abk89063 Oligonuc1
C 39	7	19.4	45	6	ABN84478	Abn84478 Oligonuc1
C 40	7	19.4	45	10	ABZ58892	Abz58892 pEt42A co
C 41	7	19.4	45	11	ADM86528	Adm86528 DNA fragm
C 42	7	19.4	50	12	ADL72800	Adl72800 siRNA-pro
C 43	7	19.4	57	6	ABN84479	Abn84479 Oligonuc1
44	7	19.4	57	10	ABZ58893	Abz58893 pEt42A co
45	7	19.4	57	11	ADM86529	Adm86529 DNA fragm

ALIGNMENTS

RESULT 1
AAX99247
ID AAX99247 standard; DNA; 627 BP.
XX
AC AAX99247;
XX
DT 04-OCT-1999 (first entry)
XX
DE Nucleotide sequence encoding a GP97ompPR fusion protein.
XX
KW Recombinant peptide production; fusion peptide; accessory peptide;
KW chemical modification; agglutination; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 82..546
FT /tag= a
XX
PN WO9938984-A1.
XX
PD 05-AUG-1999.
XX
PF 29-JAN-1999; 99WO-JP000406.
XX
PR 30-JAN-1998; 98JP-00032272.
XX
PA (SUNR) SUNTORY LTD.
XX

PI Ohsuye K, Yabuta M, Suzuki Y;
 DR WPI; 1999-469333/39.
 DR P-PSDB; AAY28950.
 XX
 XX Production of recombinant peptide as a fusion peptide followed by
 PT cleavage to give the target peptide in high yield and purity.
 FT
 XX
 PS Disclosure; Fig 8; 88pp; Japanese.
 XX
 XX The invention provides a method for the production of a desired
 CC recombinant peptide where the recombinant peptide is produced as a fusion
 CC peptide with a suitable accessory peptide. The fusion peptide is purified
 CC from culture, subjected to any desired chemical modification, and then
 CC cleaved and further purified to yield the desired peptide. The method is
 CC used for efficient production of the desired peptides in high yield and
 CC purity. The method allows the isoelectric point of the peptide purified
 CC from the culture medium to be regulated to prevent agglutination of the
 CC peptide and resultant loss of purity. Purity of above 98% and endotoxin
 CC content below 0.03 U/mg can be achieved. The present sequence represents
 CC a nucleotide sequence coding for fusion protein comprising GLP-1, a
 CC supplementary peptide and beta-galactosidase protective peptide
 XX
 SQ Sequence 627 BP; 134 A; 173 C; 163 G; 157 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.26e-05 Length: 627
 Score: 15.00 Matches: 15
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.67% Indels: 0
 DB: 2 Gaps: 0

US-10-500-018A-27 (1-36) x AAX99247 (1-627)

OY 20 ArgLeuTyrArgArgHisHisGlySerGlySerProSerArgHis 34
 DE |||||
 Db 400 CGCTGTATCGCGTATCATACGGTTCGGATCCCTTCGACAT 444

RESULT 2
 ADL66849
 ID ADL66849 standard; DNA; 36 BP.
 XX
 AC ADL66849;
 DT 20-MAY-2004 (first entry)
 DE
 XX Ghrelin and growth hormone secretagogue-related PCR primer 1.
 XX drug; Ghrelin; growth hormone secretagogue; PCR; primer; ss.
 XX Unidentified.
 XX WO2003084983-A1.
 XX
 XX 16-OCT-2003.
 XX
 XX 10-APR-2003; 2003WO-JP004590.
 XX
 XX 11-APR-2002; 2002JP-00109761.
 XX
 XX (DAII-) DAIICHI SUNTORY PHARMA CO LTD.
 PA (KANG/) KANGAWA K.
 XX
 XX Minamitake Y, Matsumoto M, Makino T;
 PI WPI; 2004-098839/10.
 DR
 XX Producing modified peptides or proteins with physiological activity
 FT comprises fusing side chain-modified peptide fragments obtained by solid-
 PT phase synthesis and non-modified peptides by genetic modification.
 XX
 XX Example 1; SEQ ID NO 28; 120pp; Japanese.

XX The invention relates to a novel method for producing protected or
 CC modified peptide fragments containing desired sequences of amino acids
 CC and/or non-amino acids, at least one of which is suitably protected for
 CC preparing the peptide fragment with use of a weakly acidic-eliminating
 CC resin to enable cleavage from the resin under weakly acidic conditions.
 CC The method of the invention may be used for producing modified peptide
 CC fragments of proteins and producing protected peptide fragments that do
 CC not contain any modified (non-)amino acids for application as drugs. The
 CC current sequence is that of a Ghrelin and growth hormone secretagogue-
 CC related PCR primer of the invention.
 XX
 SQ Sequence 36 BP; 6 A; 14 C; 9 G; 7 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.189 Length: 36
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 27.78% Indels: 0
 DB: 12 Gaps: 0

US-10-500-018A-27 (1-36) x ADL66849 (1-36)

OY 27 GlySerGlySerProSerArgHisArgArg 36
 Db 1 GGTTCGGATCCCTTCGACATCGCGG 30

RESULT 3
 AAX99259/C
 ID AAX99259 standard; DNA; 55 BP.
 XX
 XX AAX99259;
 AC
 XX
 DT 04-OCT-1999 (first entry)
 DE
 XX Nucleotide sequence encoding a peptide containing a site cleaved by Ompr.
 XX Recombinant peptide production; fusion peptide; accessory peptide;
 KW chemical modification; agglutination; Ompr; ss.
 XX Synthetic.
 XX WO9938984-A1.
 PN
 PD 05-AUG-1999.
 XX
 XX 29-JAN-1999; 95WO-JP000406.
 XX
 XX 30-JAN-1998; 98JP-00032272.
 XX
 XX (SUNR) SUNTORY LTD.
 PA
 XX Ohsuye K, Yabuta M, Suzuki Y;
 PI WPI; 1999-469333/39.
 DR
 XX Production of recombinant peptide as a fusion peptide followed by
 PT cleavage to give the target peptide in high yield and purity.
 XX
 PS Disclosure; Fig 6; 88pp; Japanese.
 XX
 XX The invention provides a method for the production of a desired
 CC recombinant peptide where the recombinant peptide is produced as a fusion
 CC peptide with a suitable accessory peptide. The fusion peptide is purified
 CC from culture, subjected to any desired chemical modification, and then
 CC cleaved and further purified to yield the desired peptide. The method is
 CC used for efficient production of the desired peptides in high yield and
 CC purity. The method allows the isoelectric point of the peptide purified
 CC from the culture medium to be regulated to prevent agglutination of the
 CC peptide and resultant loss of purity. Purity of above 98% and endotoxin
 CC content below 0.03 U/mg can be achieved. The present sequence represents
 CC a nucleotide sequence coding for an amino acid sequence containing a site

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 12:00:38 ; Search time 1763 Seconds
(without alignments)
989.443 Million cell updates/sec

Title: US-10-500-018a-27

Perfect score: 36

Sequence: 1 BPHHHPGGRQMGYDADVRLYRRHHGSGSPSRHR 36

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9405349

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xl
-Q/cgn2.1/USPTO.spool.p/US10500018/runat.25032005.092649.24493/app.query.fasta.1.199
-DB=GenEmbl -QFT=fastap -SUPFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10500018 @CGN 1.1.5600 @runat.25032005.092649.24493 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRAIDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	10	27.8	224671	2	AC120806
2	10	27.8	230196	2	AC125942
3	9	25.0	4589	10	BC056370
4	9	25.0	110000	1	AP006840_19
					Continuation (20 o

5	9	25.0	191247	9	AL645944
6	9	25.0	196812	10	AL645615
7	9	25.0	294800	1	SME591789
c 8	8	22.2	292	6	AR358347
c 9	8	22.2	292	6	AR359903
c 10	8	22.2	304	1	EC0302767
c 11	8	22.2	345	1	KSP302771
c 12	8	22.2	1284	6	BD180545
c 13	8	22.2	1950	6	BD180542
c 14	8	22.2	2111	6	A64844
c 15	8	22.2	2374	8	AK101369
c 16	8	22.2	3277	3	PLY11145
c 17	8	22.2	3282	6	CQ363864
c 18	8	22.2	7914	1	AF021262
c 19	8	22.2	7950	1	AF016298
c 20	8	22.2	8674	1	AE012354
c 21	8	22.2	9623	6	AR354048
c 22	8	22.2	9623	6	AR355604
c 23	8	22.2	12373	1	PSF17897
c 24	8	22.2	13045	1	AE004906
c 25	8	22.2	13542	1	AE005090
c 26	8	22.2	13842	1	AF247502
c 27	8	22.2	18030	1	AY249242
c 28	8	22.2	20103	1	STAF000001
c 29	8	22.2	37175	2	AC144571
c 30	8	22.2	49898	1	AC027136
c 31	8	22.2	62423	3	AF004554
c 32	8	22.2	69431	3	AF067612
c 33	8	22.2	74269	9	AL355865
c 34	8	22.2	74845	9	AL355865
c 35	8	22.2	82884	3	AC084439
c 36	8	22.2	86158	8	AP004013
c 37	8	22.2	94779	2	AC151672
c 38	8	22.2	95802	1	EX571857_27
c 39	8	22.2	100986	8	AC105363
c 40	8	22.2	105521	8	AP005485
c 41	8	22.2	107696	2	AC150197
c 42	8	22.2	109528	1	AP003366
c 43	8	22.2	110000	1	AE017180_34
c 44	8	22.2	110000	1	AE017283_10
c 45	8	22.2	110000	1	AP006840_17

ALIGNMENTS

RESULT 1	AC120806	224671 bp	DNA	linear	HTG 21-SEP-2002
AC120806/c	Rattus norvegicus clone CH230-23K3, ***	SEQUENCING IN PROGRESS ***			
LOCUS	2 unordered pieces.				
DEFINITION	AC120806	GI:23265334			
ACCESSION	AC120806.3	GI:23265334			
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.				
KEYWORDS	Rattus norvegicus (Norway rat)				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
REFERENCE	1 (bases 1 to 224671)				
AUTHORS	Munzy,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmad,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J.J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,				

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregregis, E., Geer, K., Gilly, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Huliyk, S., Hume, J., Idlebirt, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Li, Z., Liu, J., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheeshwari, M., Mahindratne, M., Mahmoud, R., Mallory, K., Mangum, A., Mangum, B., Napua, P., Martin, K., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Morgan, M., Morris, K., Morris, S., Munidasa, M., Nguyen, N., Norris, S., Nankervis, C., Neal, D., Newton, N., Olanpunaagoo, A., Pal, S., Parks, K., Nwalele, O., Okwuonu, G., Olanpunaagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Pioppner, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Rose, R., Ruiz, S.J., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Richards, S., Riggs, F., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyan, E., Sisson, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Unpublished
Direct Submission
2 (bases 1 to 224671)
Worley, K.C.
Direct Submission
Submitted (10-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 224671)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21908080.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKYS
Center clone name: CH230-23K3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 205843 bases at least Q40
Consensus quality: 208076 bases at least Q30
Consensus quality: 209636 bases at least Q20
Estimated insert size: 227577; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a working draft sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 222786: contig of 222786 bp in length
* 222787 222886: gap of unknown length
* 222887 224671: contig of 1785 bp in length.

FEATURES

source

Location/Qualifiers

1..224671

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-23K3"

misc_feature

1..1309

/note="wgs end extension"

clone_end:Sp6"

4908..6155

/note="wgs end extension"

clone_end:Sp6"

6812..9311

/note="wgs end extension"

clone_end:Sp6"

complement(9377..10223)

/note="clone boundary"

clone_end:Sp6"

site:EcoRI

end sequence:BH274476"

complement(213906..214665)

/note="clone boundary"

clone_end:T7"

site:EcoRI

end sequence:BH274475"

215673..217452

/note="wgs end extension"

clone_end:T7"

221188..222786

/note="wgs end extension"

clone_end:T7"

misc_feature

201

Length: 224671

Matches: 10

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-500-018A-27 (1-36) x AC120806 (1-224671)

QY 23 ArgArgHisGlySerGlySerProSer 32

DB 38125 CGGAGCACCACGATCCGATCTCCTCT 38096

RESULT 2

AC125942

LOCUS

DEFINITION Rattus norvegicus clone CH230-10117, *** SEQUENCING IN PROGRESS

230196 bp DNA linear HTG 13-NOV-2002

AC125942

VERSION

AC125942.4 GI:24940667

KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: March 26, 2005, 11:08:33 ; Search time 495 Seconds
(without alignments)
433.406 Million cell updates/sec
Title: US-10-500-018A-27
Perfect score: 215
Sequence: 1 EPHHHPGQRQMGYDADVRLYRRHHGSGSPSRHRR 36
Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0
Searched: 5532208 seqs, 2979665951 residues
Total number of hits satisfying chosen parameters: 11104416
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlsp
-Q/cgn2_1/USPTO_spool_p/US10500018/runat_25032005_092624_24278/app_query.fasta_1.199
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USFR=US10500018 @CGN 1.1 723 @runat_25032005_092624_24278
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
C 1	84	39.1	339	15	US-10-156-761-2150
2	84	39.1	9025608	15	US-10-156-761-1
3	76.5	35.6	1635	18	US-10-437-963-45784
C 4	76	35.3	868	18	US-10-437-963-101122
5	76	35.3	1539	17	US-10-355-238-12
6	75.5	35.1	819	17	US-10-369-493-42919
7	75	34.9	1136	18	US-10-437-963-24983
C 8	75	34.9	3138	15	US-10-156-761-2538
9	74	34.4	1320	15	US-10-156-761-5573
10	73.5	34.2	1280	17	US-10-435-114-17735
11	72.5	33.7	1129	18	US-10-437-963-75929
C 12	72	33.5	1278	17	US-10-369-493-42314
13	72	33.5	17710	9	US-09-956-004-70
14	72	33.5	17710	18	US-10-808-570-70
15	72	33.5	23654	16	US-10-085-959-11
16	71.5	33.3	263	18	US-10-425-115-120427
17	71.5	33.3	840	15	US-10-156-761-4830
C 18	71.5	33.3	2833	15	US-10-190-471-5
19	71	33.0	513	9	US-09-864-761-13346
C 20	71	33.0	1080	18	US-10-437-963-23219
21	71	33.0	1449	15	US-10-156-761-6862
22	71	33.0	1881	15	US-10-156-761-1804
23	71	33.0	2082	17	US-10-369-493-41602
24	71	33.0	8906	17	US-10-374-979-85
25	71	33.0	8906	17	US-10-331-496A-87
26	71	33.0	8906	17	US-10-182-936A-85
27	71	33.0	8906	18	US-10-477-238A-664
C 28	71	33.0	8906	18	US-10-680-287A-664
29	71	33.0	9025608	15	US-10-156-761-1
30	70.5	32.8	901	17	US-10-369-493-31770
C 31	70.5	32.8	1531	18	US-10-437-963-79838
32	70.5	32.8	1758	18	US-10-411-910A-102
C 33	70.5	32.8	1882	18	US-10-437-963-102401
34	70.5	32.8	2100	17	US-10-282-122A-12927
35	70.5	32.8	3213	17	US-10-369-493-42962
36	70	32.6	1386	17	US-10-375-932-212
37	70	32.6	1596	17	US-10-282-122A-30174
38	70	32.6	2043	17	US-10-375
C 39	70	32.6	2406	17	US-10-282-122A-14358
40	70	32.6	2577	18	US-10-425-115-57282
41	70	32.6	2988	18	US-10-437-963-78998
42	70	32.6	4434	17	US-10-369-493-26978
43	69.5	32.3	1229	18	US-10-425-115-12306
44	69.5	32.3	1266	18	US-10-411-910A-95
45	69.5	32.3	1773	17	US-10-282-122A-14926

ALIGNMENTS

RESULT 1
US-10-156-761-2150/c
; Sequence 2150, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, KASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2150
; LENGTH: 339
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(339)
US-10-156-761-2150
Alignment Scores:
Pred. No.: 0.0238 Length: 339
Score: 84.00 Matches: 15
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 55.56% Mismatches: 9
Query Match: 39.07% Indels: 0
DB: 15 Gaps: 0
US-10-500-018A-27 (1-36) x US-10-156-761-2150 (1-339)
Qy 1 GluProHisHisHisProGlyGlyArgGlnMethHisGlyTyrAspAlaAaspValArg 20
Db 141 GACCCCCACACGAGAGCGCCATTGGACGGCGGTTCACGGCTTCGACGCCCGAGGGTGC 82
Qy 21 LeuTyrArgArgHisHisGly 27
Db 81 CTTGCCCGCGCGGACGACGCG 61
RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 1.07e+03 Length: 9025608
Score: 84.00 Matches: 15
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 55.56% Mismatches: 9
Query Match: 39.07% Indels: 0
DB: 15 Gaps: 0
US-10-500-018A-27 (1-36) x US-10-156-761-1 (1-9025608)
Qy 1 GluProHisHisHisHisProGlyGlyArgGlnMethHisGlyTyrAspAlaAaspValArg 20
Db 2634908 GACCCCCACACGAGAGCGCCATTGGACGGCGGTTCACGGCTTCGACGCCCGAGGGTGC 2634967
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2150
; LENGTH: 339
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(339)
US-10-156-761-2150
Alignment Scores:
Pred. No.: 0.0238 Length: 339
Score: 84.00 Matches: 15
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 55.56% Mismatches: 9
Query Match: 39.07% Indels: 0
DB: 15 Gaps: 0
US-10-500-018A-27 (1-36) x US-10-156-761-2150 (1-339)
Qy 1 GluProHisHisHisProGlyGlyArgGlnMethHisGlyTyrAspAlaAaspValArg 20
Db 141 GACCCCCACACGAGAGCGCCATTGGACGGCGGTTCACGGCTTCGACGCCCGAGGGTGC 82
Qy 21 LeuTyrArgArgHisHisGly 27
Db 81 CTTGCCCGCGCGGACGACGCG 61
RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 1.07e+03 Length: 9025608
Score: 84.00 Matches: 15
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 55.56% Mismatches: 9
Query Match: 39.07% Indels: 0
DB: 15 Gaps: 0
US-10-500-018A-27 (1-36) x US-10-156-761-1 (1-9025608)
Qy 1 GluProHisHisHisHisProGlyGlyArgGlnMethHisGlyTyrAspAlaAaspValArg 20
Db 2634908 GACCCCCACACGAGAGCGCCATTGGACGGCGGTTCACGGCTTCGACGCCCGAGGGTGC 2634967

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 10:20:57 ; Search time 132 Seconds
(without alignments)
446.257 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 215

Sequence: 1 EPHHHPGQRMHGYDADVRLYRRHGGSPSRHR 36

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DRV=xl

-Q=/cgn2_1/USFTO.spool_P/US10500018/runat_25032005_092623_24191/app_query.fasta_1.199

-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10500018 @CGN 1.1 105 @runat_25032005_092623_24191 -NCPU=6 -ICPU=3

-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG

-DRV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCUTUS_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82.5	38.4	2277	4	US-09-902-540-2843
2	82.5	38.4	16047	4	US-09-902-540-1136
3	80.5	37.4	3123	4	US-09-902-540-3802
4	80.5	37.4	9567	4	US-09-902-540-1002
5	75.5	35.1	843	4	US-09-902-540-8507
6	75.5	35.1	6482	4	US-09-902-540-896
7	73.5	34.2	693	4	US-09-902-540-8474
8	73.5	34.2	7513	4	US-09-902-540-892
9	73	34.0	6828	4	US-09-902-540-8354
10	73	34.0	8056	4	US-09-902-540-874
11	73	34.0	9729	4	US-09-902-540-948
12	72	33.5	429	4	US-09-252-991A-8302

C	13	72	33.5	1395	4	US-09-252-991A-8312	Sequence 8312, Ap
	14	72	33.5	1398	4	US-09-252-991A-8277	Sequence 8277, Ap
	15	72	33.5	1608	4	US-09-252-991A-8157	Sequence 8157, Ap
C	16	72	33.5	1734	4	US-09-252-991A-7986	Sequence 7986, Ap
	17	72	33.5	17710	3	US-08-976-259-70	Sequence 70, Appl
	18	72	33.5	17710	4	US-09-956-004-70	Sequence 70, Appl
	19	71	33.0	8865	4	US-09-949-016-4238	Sequence 4238, Ap
	20	71	33.0	8906	2	US-08-826-267-1	Sequence 1, Appl1
	21	71	33.0	239527	4	US-09-949-016-15980	Sequence 15980, A
	22	70.5	32.8	3252	4	US-09-902-540-9254	Sequence 9254, Ap
	23	70.5	32.8	10322	4	US-09-902-540-989	Sequence 989, App
	24	69	32.1	1248	4	US-09-902-540-2492	Sequence 2492, Ap
	25	69	32.1	1452	4	US-09-252-991A-4444	Sequence 4444, Ap
	26	69	32.1	1671	4	US-09-252-991A-4648	Sequence 4648, Ap
C	27	69	32.1	2115	4	US-09-252-991A-4163	Sequence 4163, Ap
	28	69	32.1	2403	4	US-09-902-540-4394	Sequence 4394, Ap
C	29	69	32.1	17938	4	US-09-902-540-1111	Sequence 1111, Ap
	30	69	32.1	29272	4	US-09-902-540-1217	Sequence 1217, Ap
C	31	68.5	31.9	828	4	US-09-252-991A-15716	Sequence 15716, A
	32	68.5	31.9	899	3	US-09-038-542-1	Sequence 1, Appl1
	33	68.5	31.9	2211	4	US-09-252-991A-15685	Sequence 15685, A
C	34	68	31.6	1017	4	US-09-902-540-4321	Sequence 4321, Ap
	35	68	31.6	1294	3	US-09-312-038-4	Sequence 4, Appl1
	36	68	31.6	1294	4	US-09-850-964-4	Sequence 4, Appl1
	37	68	31.6	2289	3	US-09-312-038-3	Sequence 3, Appl1
	38	68	31.6	2289	4	US-09-850-964-3	Sequence 3, Appl1
C	39	68	31.6	25733	4	US-09-902-540-1215	Sequence 1215, Ap
	40	67.5	31.4	717	4	US-09-486-241-31	Sequence 31, Appl
	41	67.5	31.4	1074	4	US-09-252-991A-13570	Sequence 13570, A
	42	67.5	31.4	2673	4	US-09-252-991A-2993	Sequence 2993, Ap
C	43	67.5	31.4	2715	4	US-09-252-991A-2686	Sequence 2686, Ap
	44	67.5	31.4	2814	4	US-09-252-991A-2875	Sequence 2875, Ap
C	45	67.5	31.4	3129	4	US-09-252-991A-13873	Sequence 13873, A

ALIGNMENTS

RESULT 1

US-09-902-540-2843

; Sequence 2843, Application US/0902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15849)B

; CURRENT APPLICATION NUMBER: US/09/902,540

; CURRENT FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883

; PRIOR FILING DATE: 2000-07-10

; NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 2843

; LENGTH: 2277

; TYPE: DNA

; ORGANISM: Myxococcus xanthus

US-09-902-540-2843

Alignment Scores:

Pred. No.: 2.2 Length: 2277

Score: 82.50 Matches: 17

Percent Similarity: 58.06% Conservative: 1

Best Local Similarity: 54.84% Mismatches: 12

Query Match: 38.37% Indels: 1

DB: 4 Gaps: 1

US-10-500-018A-27 (1-36) x US-09-902-540-2843 (1-2277)

QY 2 ProHisHisHisProGlyArgGlnMetHisGlyTyzAspAlaAspValArg--- 20

DB 1359 CCACACCTCCATCACCTCGAAGAGGCCCAACATCATCCAGATGACCGCGGACAG 1418

```

QY      21 LeuTyrArgArgHisHisGlySerGlySerPro 31
DB      1419 CTTGGGGGGCTCACCACCAAGCAGGTCAACCT 1451

RESULT 2
US-09-902-540-1136/c
; Sequence 1136, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1136
; LENGTH: 16047
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1136

Alignment Scores:
Pred. No.:      19.4      Length:      16047
Score:          82.50     Matches:      17
Percent Similarity: 58.06% Conservative: 1
Best Local Similarity: 54.84% Mismatches: 12
Query Match:    38.37% Indels:      1
DB:             4        Gaps:       1

US-10-500-018A-27 (1-36) x US-09-902-540-1136 (1-16047)

QY      2 ProHisHisHisProGlyArgGlnMetHisGlyTyrAspAlaAspValArg--- 20
DB      11382 CCACACTTCATCACCCTGGAGAGGCCCCCAACATCATCCAGATGACCGCGGAGACAG 11323

QY      21 LeuTyrArgArgHisHisGlySerGlySerPro 31
DB      11322 CTTGGGGGGCTCACCACCAAGCAGGTCAACCT 11290

RESULT 3
US-09-902-540-3802
; Sequence 3802, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3802
; LENGTH: 3123
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3802

Alignment Scores:
Pred. No.:      5.32      Length:      3123
Score:          80.50     Matches:      18
Percent Similarity: 42.86% Conservative: 6
Best Local Similarity: 32.14% Mismatches: 9
Query Match:    37.44% Indels:      23
DB:             4        Gaps:       2

```

```

US-10-500-018A-27 (1-36) x US-09-902-540-3802 (1-3123)

QY      3 HisHisHisHisProGlyArgGlnMetHisGlyTyr----- 15
DB      51 CATCATCATCACCCTGGTGGGGCCATCTCCATCCCATCTGCGCCATCGAGCAATACCC 110

QY      16 -----AspAlaAspVal----- 19
DB      111 TGACCTGTGTCGCCCGGTGGTGAGGTACCGCGACGTACATCGCGCGCTCCGCGGAGAC 170

QY      20 ArgLeuTyrArgArgHisHisGlySerGlySerProSerArgHisArg 35
DB      171 CGTGGAGAGCGCGCTCACCCCGTCTGGAGCGCAGCTCAACGCGCT 218

RESULT 4
US-09-902-540-1002
; Sequence 1002, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1002
; LENGTH: 9567
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1002

Alignment Scores:
Pred. No.:      18.5      Length:      9567
Score:          80.50     Matches:      18
Percent Similarity: 42.86% Conservative: 6
Best Local Similarity: 32.14% Mismatches: 9
Query Match:    37.44% Indels:      23
DB:             4        Gaps:       2

US-10-500-018A-27 (1-36) x US-09-902-540-1002 (1-9567)

QY      3 HisHisHisHisProGlyArgGlnMetHisGlyTyr----- 15
DB      1238 CATCATCATCACCCTGGTGGGGCCATCTCCATCCCATCTGCGCCATCGAGCAATACCC 1297

QY      16 -----AspAlaAspVal----- 19
DB      1298 TGACCTGTGTCGCCCGGTGGTGAGGTACCGCGACGTACATCGCGCGCTCCGCGGAGAC 1357

QY      20 ArgLeuTyrArgArgHisHisGlySerGlySerProSerArgHisArg 35
DB      1358 CGTGGAGAGCGCGCTCACCCCGTCTGGAGCGCAGCTCAACGCGCT 1405

RESULT 5
US-09-902-540-8507
; Sequence 8507, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 10:14:07 ; Search time 3094 Seconds
(without alignments)
442.894 Million cell updates/sec

Title: US-10-500-018a-27

Perfect score: 215

Sequence: 1 EPHHHPGQMGYDADVRLYRRHHGSGPSRHR 36

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10500018/runat_25032005_092623_24172/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10500018 @CGN 1 1 5180 @runat_25032005_092623_24172 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -HEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hgc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	81.5	37.9	632	CA368943	CA368943 645280 NC
C 2	79.5	37.0	458	CO549635	CO549635 LYEST7304
C 3	79.5	37.0	641	CO547916	CO547916 LYEST4584
C 4	78.5	36.5	491	BH869043	BH869043 hl41f08.b
C 5	78.5	36.5	656	CG149946	CG149946 PUUCV19TD
C 6	78.5	36.5	658	CG149944	CG149944 PUUCV19TD
C 7	78.5	36.5	675	CC714012	CC714012 OGUUN74TV
C 8	78.5	36.5	720	CC333096	CC333096 OGMET52TV
C 9	78.5	36.5	876	CG268453	CG268453 OGWGP77TH

C 10	78.5	36.5	880	9	CG370812	CG370812 OGYCK51TH
C 11	78.5	36.5	903	9	CG370826	CG370826 OGYCK51TV
C 12	78.5	36.5	907	8	BZ805398	PUGBC01TD
C 13	78.5	36.5	926	9	CC598158	OSVAC21TC
C 14	78.5	36.5	1021	9	CG220582	OGK054TV
C 15	78.5	36.5	1023	8	BZ805394	PUGBC01TB
C 16	78	36.3	680	5	BU377778	BZ805394
C 17	77	35.8	564	5	BU646799	1112053P0
C 18	77	35.8	750	9	CC822438	MBSP15G11
C 19	77	35.8	835	8	CC463846	ZMMBB038
C 20	77	35.8	986	8	CC463800	ZMMBB038
C 21	76.5	35.6	317	1	AL830450	AL830450
C 22	76.5	35.6	594	7	CN011399	WHE3883 G
C 23	76.5	35.6	647	8	BZ691280	SP_Ba000
C 24	76.5	35.6	698	4	BZ288659	BZ691280
C 25	76.5	35.6	714	4	BZ294819	BZ288659
C 26	76.5	35.6	848	6	CB960452	AGENCOURT
C 27	75.5	35.1	139	7	CN641047	291H05_55
C 28	75.5	35.1	357	8	BH787215	fzmb014f0
C 29	75.5	35.1	715	9	CW008390	ZMMLA001
C 30	75	34.9	480	6	CD444601	EL01N0441
C 31	75	34.9	518	7	CK368933	ZMTW0055
C 32	75	34.9	559	8	AQ421625	RPCI-11-1
C 33	75	34.9	604	8	AQ375679	RPCI11-16
C 34	75	34.9	798	9	CG734320	RP11-160B
C 35	75	34.9	1001	4	BG333683	602460574
C 36	75	34.9	1047	9	CL969959	OGIFCC019
C 37	74.5	34.7	645	4	BZ518816	BZ518816
C 38	74.5	34.7	648	4	BM595281	BM595281
C 39	74.5	34.7	667	4	BZ531360	BZ531360
C 40	74.5	34.7	688	4	BZ028582	BZ028582
C 41	74.5	34.7	688	4	BM621668	BM621668
C 42	74.5	34.7	690	4	BZ541561	BZ541561
C 43	74.5	34.7	737	4	BZ723996	BZ723996
C 44	74.5	34.7	739	4	BZ723991	BZ723991
C 45	74.5	34.7	801	9	CG341075	OGWLI267TH

ALIGNMENTS

RESULT 1
CA368943/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CA368943 645280 NCCWA 1RT Oncorhynchus mykiss cdna linear EST 06-NOV-2002
632 bp mRNA
645280 NCCWA 1RT Oncorhynchus mykiss cdna clone lRT10H12_D_D06 5',
mRNA sequence.

CA368943

CA368943.1 GI:24680670

EST.

Oncorhynchus mykiss (rainbow trout)

Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 632)

Rexroad, C.B. 3rd, Lee, Y., Keele, J.W., Karamycheva, S., Brown, G.,

Koop, B., Gahr, S.A., Palti, Y. and Quackenbush, J.

Sequence analysis of a rainbow trout cdna library and creation of a

gene index

Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)

Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by

cross match v0.990329.

Seq primer: AGCGGATACAAATTCACACAGGA.

Location/Qualifiers

1..632

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"
 /clone="INT10H12 D_D06"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="NCCGWA 1RT"
 /notes="vector: pCMV SPOR6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."

ORIGIN

Alignment Scores:
 Pred. No.: 20.6 Length: 632
 Score: 81.50 Matches: 17
 Percent Similarity: 58.82% Conservative: 3
 Best Local Similarity: 50.00% Mismatches: 9
 Query Match: 37.91% Indels: 5
 DB: 6 Gaps: 2

US-10-500-018A-27 (1-36) x CA368943 (1-632)

QY 2 ProHisHisHisProGlyArgGlnMethHisGlyTyrAspAlaaspValArgLeu 21
 |||||
 Db 598 CCCATCATCATCAT-----CGTCGTCGTGTTGGG-----GCGGACACCCATGCC 554
 |||||

QY 22 TyrArgHisHisGlySerGlySerProSerArgHisArg 35
 :|||

Db 553 TTGCAGAGACATCATCATGTTGGACACCCCTCAGGCACAGG 512
 |||||

RESULT 2

CO549635/c
 LOCUS 20.6 bp mRNA linear EST 01-SEP-2004
 DEFINITION LYEST7304 Sea lamprey LyEST Petromyzon marinus cDNA, mRNA sequence.
 ACCESSION CO549635
 VERSION CO549635.1 GI:51797951
 KEYWORDS EST.
 SOURCE Petromyzon marinus (sea lamprey)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 Petromyzontiformes; Petromyzontidae; Petromyzon.

REFERENCE
 1 (bases 1 to 458)
 Pancer, Z., Mayer, W.E., Klein, J. and Cooper, M.D.

Prototypic T-cell receptor and CD4-like coreceptor expressed in
 lymphocytes of the agnathan sea lamprey

Proc. Natl. Acad. Sci. U.S.A. 101, 13273-13278 (2004)

JOURNAL

COMMENT
 Contact: Pancer, Zeev
 Division of Developmental and Clinical Immunology
 The University of Alabama at Birmingham
 378 Wallace Tumor Institute, 1530 Third Avenue, South, Birmingham,
 AL 35294-3300

Tel: 205-975-5812
 Fax: 205-975-7218
 Email: zpancer@uab.edu.

Location/Qualifiers

FEATURES

source

1..458
 /organism="Petromyzon marinus"
 /mol_type="mRNA"
 /db_xref="taxon:7757"
 /cell_type="lymphocyte"
 /dev_stage="unstimulated larvae"
 /clone_lib="Sea lamprey LyEST"
 /note="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs
 from unstimulated larvae. All are from arrayed colonies
 from a directionally cloned cDNA library in Lambda ZAP
 Express (Stratagene). All are single pass 5' sequences."

ORIGIN

Alignment Scores:
 Pred. No.: 24.7 Length: 458
 Score: 79.50 Matches: 16
 Percent Similarity: 44.19% Conservative: 3
 Best Local Similarity: 37.21% Mismatches: 15
 Query Match: 36.98% Indels: 9
 DB: 7 Gaps: 1

US-10-500-018A-27 (1-36) x CO549635 (1-458)

QY 2 ProHisHisHisHisProGlyArgGlnMethHisGlyTyrAspAlaaspValArgLeu 21
 |||||

Db 361 CCACACGGACACACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGACAC 302
 |||||

QY 22 Tyr-----ArgArgHisHisGlySerGlySerProSer 32
 :|||

Db 301 CACACGGACACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGCT 242
 |||||

QY 33 ArgHisArg 35
 |||||

Db 241 GCCCACCAG 233
 |||||

RESULT 3

CO547916/c

LOCUS 641 bp mRNA linear EST 01-SEP-2004
 DEFINITION LYEST4584 Sea lamprey LyEST Petromyzon marinus cDNA, mRNA sequence.
 ACCESSION CO547916
 VERSION CO547916.1 GI:51796228
 KEYWORDS EST.
 SOURCE Petromyzon marinus (sea lamprey)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 Petromyzontiformes; Petromyzontidae; Petromyzon.

REFERENCE
 1 (bases 1 to 641)
 Pancer, Z., Mayer, W.E., Klein, J. and Cooper, M.D.

Prototypic T-cell receptor and CD4-like coreceptor expressed in
 lymphocytes of the agnathan sea lamprey

Proc. Natl. Acad. Sci. U.S.A. 101, 13273-13278 (2004)

JOURNAL

COMMENT

Contact: Pancer, Zeev
 Division of Developmental and Clinical Immunology
 The University of Alabama at Birmingham
 378 Wallace Tumor Institute, 1530 Third Avenue, South, Birmingham,
 AL 35294-3300

Tel: 205-975-5812
 Fax: 205-975-7218
 Email: zpancer@uab.edu.

Location/Qualifiers

FEATURES

source

1..641
 /organism="Petromyzon marinus"
 /mol_type="mRNA"
 /db_xref="taxon:7757"
 /cell_type="lymphocyte"
 /dev_stage="unstimulated larvae"
 /clone_lib="Sea lamprey LyEST"
 /note="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs
 from unstimulated larvae. All are from arrayed colonies
 from a directionally cloned cDNA library in Lambda ZAP
 Express (Stratagene). All are single pass 5' sequences."

ORIGIN

Alignment Scores:
 Pred. No.: 35.4 Length: 641
 Score: 79.50 Matches: 16
 Percent Similarity: 44.19% Conservative: 3
 Best Local Similarity: 37.21% Mismatches: 15
 Query Match: 36.98% Indels: 9
 DB: 7 Gaps: 1

US-10-500-018A-27 (1-36) x CO547916 (1-641)

QY 2 ProHisHisHisHisProGlyArgGlnMethHisGlyTyrAspAlaaspValArgLeu 21
 |||||

Db 463 CCACACGGACACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGACAC 404
 |||||

QY 22 Tyr-----ArgArgHisHisGlySerGlySerProSer 32
 :|||

Db 403 CACACGGACACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGACACCGGCT 344
 |||||

QY 33 ArgHisArg 35
 |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 09:07:42 ; Search time 431 Seconds
(without alignments)
494.456 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 215

Sequence: 1 EPHHHPGGRQMGGYDADVRLYRRHHGSGSPSRHR 36

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlp
-Q/cgn2_1/USPTO spool_p/US10500018/runat_25032005_092622_24150/app_query.fasta_1.199
-DB=N_Geneseq_16Dec04 -QFWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10500018 @CGN 1.1 708 @runat_25032005_092622_24150 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- N Geneseq_16Dec04:*
- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2002as:*
- 6: geneseqn2002bs:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2003ds:*
- 11: geneseqn2004as:*
- 12: geneseqn2004bs:*
- 13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	59.5	627	2 AAX99247	Aax99247 Nucleotid
c 2	92	42.8	55	2 AAX99259	Aax99259 Nucleotid
c 3	92	42.8	55	3 AAA99157	Aaa99157 Ompt prot
4	87	40.5	69	2 AAX99241	Aax99241 Nucleotid
5	87	40.5	69	3 AAA99154	Aaa99154 Ompt prot

c	6	87	40.5	69	3	AAA99155	Aaa99155 Ompt prot
c	8	81	37.7	2391	4	ABL14997	Ab114997 Drosophil
c	9	81	37.7	6629	4	ABL14996	Ab114996 Drosophil
c	9	78	36.3	47	2	AAX99242	Aax99242 Nucleotid
c	10	78	36.3	47	2	AAX99156	Aaa99156 Ompt prot
c	11	75.5	35.1	819	13	ADT44481	Adt44481 Bacterial
c	12	75	34.9	1788	10	ACF04832	Acf04832 M licheni
c	13	75	34.9	51855	10	ACF04818	Acf04818 Melithiaz
c	14	74	34.4	6702	12	ADE43905	Adg43905 Human DP4
c	15	74	34.4	8010	10	ADG42110	Adg42110 Human bra
c	16	72	33.5	429	11	ABD09698	Abd09698 Pseudomon
c	17	72	33.5	1278	13	ADT43876	Adt43876 Bacterial
c	18	72	33.5	1395	11	ABD09708	Abd09708 Pseudomon
c	19	72	33.5	1398	11	ABD09673	Abd09673 Pseudomon
c	20	72	33.5	1608	11	ABD09553	Abd09553 Pseudomon
c	21	72	33.5	1734	11	ABD09382	Abd09382 Pseudomon
c	22	72	33.5	2055	13	ADR38459	Adr38459 Streptomy
c	23	72	33.5	17710	2	AAV31256	Aav31256 E. coli J
c	24	72	33.5	23654	6	ABS78844	Abs78844 E. coli C
c	25	72	33.5	23654	10	ADH80411	Adh80411 Escherich
c	26	71.5	33.3	2833	10	ADC59411	Adc59411 P. pneumo
c	27	71	33.0	513	4	AAK11215	Aak11215 Human bra
c	28	71	33.0	513	6	ABS11018	Abs11018 Human gen
c	29	71	33.0	1824	12	ADM99178	Adm99178 Environme
c	30	71	33.0	2082	13	ADT43164	Adt43164 Bacterial
c	31	71	33.0	4066	12	ADQ67377	Adq67377 Novel hum
c	32	71	33.0	8586	8	ACA57512	Aca57512 Human adi
c	33	71	33.0	8906	2	AAT85392	Aat85392 Human TRI
c	34	71	33.0	8906	5	AAS81211	Aas81211 DNA encod
c	35	71	33.0	8906	5	ABA82699	Abab2699 GTPase ge
c	36	71	33.0	8906	8	ACC46019	Acc46019 Human GTP
c	37	71	33.0	8906	10	ADB98713	Adb98713 Human GTP
c	38	71	33.0	8906	10	ADB82509	Adb82509 Human DNA
c	39	71	33.0	8906	10	ADD85083	Adg85083 Encoding
c	40	71	33.0	8906	10	ADG14305	Adg14305 Human TRI
c	41	71	33.0	8908	5	AAS70499	Aas70499 DNA encod
c	42	71	33.0	9437	5	AAS81212	Aas81212 DNA encod
c	43	71	33.0	9449	5	AAS70498	Aas70498 DNA encod
c	44	71	33.0	9449	8	ACD13399	Acd13399 Human DNA
c	45	71	33.0	9450	5	AAS84943	Aas84943 DNA encod

ALIGNMENTS

RESULT 1
AAX99247
ID AAX99247 standard; DNA; 627 BP.
XX
AC AAX99247;
DT 04-OCT-1999 (first entry)
XX
DE Nucleotide sequence encoding a GP97ompPR fusion protein.
XX
KW Recombinant peptide production; fusion peptide; accessory peptide;
KW chemical modification; agglutination; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 82..546
FT /*tag= a
XX
FN WO9938984-A1.
XX
PD 05-AUG-1999.
XX
PF 29-JAN-1999; 99WO-JP000406.
XX
PR 30-JAN-1998; 98JP-00032272.
XX
PA (SUNR) SUNTORY LTD.
XX

```

PI Ohsuye K, Yabuta M, Suzuki Y;
XX WPI; 1999-469333/39.
DR P-PSDB; AAY28950.
XX
XX Production of recombinant peptide as a fusion peptide followed by
PT cleavage to give the target peptide in high yield and purity.
XX
XX Disclosure; Fig 8; 89pp; Japanese.
XX
XX The invention provides a method for the production of a desired
CC recombinant peptide where the recombinant peptide is produced as a fusion
CC peptide with a suitable accessory peptide. The fusion peptide is purified
CC from culture, subjected to any desired chemical modification, and then
CC cleaved and further purified to yield the desired peptide. The method is
CC used for efficient production of the desired peptides in high yield and
CC purity. The method allows the isoelectric point of the peptide purified
CC from the culture medium to be regulated to prevent agglutination of the
CC peptide and resultant loss of purity. Purity of above 98% and endotoxin
CC content below 0.03 U/mg can be achieved. The present sequence represents
CC a nucleotide sequence coding for fusion protein comprising GLP-1, a
CC supplementary peptide and beta-galactosidase protective peptide
XX
XX Sequence 627 BP; 134 A; 173 C; 163 G; 157 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.1e-05 Length: 627
Score: 128.00 Matches: 22
Percent Similarity: 96.00% Conservative: 2
Best Local Similarity: 88.00% Mismatches: 1
Query Match: 59.53% Indels: 0
DB: 2 Gaps: 0
US-10-500-018a-27 (1-36) x AAX99247 (1-627)
QY 12 MetHisGlyTyrAspAlaAspValArgLeuTyrArgArgHisHisGlySerGlySerPro 31
DB 376 ATGCATGGTTARGCGGAGCTCGCTGTATCGCTCATCACGGTTCGGATCCCT 435
QY 32 SerArgHisArgArg 36
DB 436 TCTCGACATCCGCGG 450
RESULT 2
AAX99259/c
ID AAX99259 standard; DNA; 55 BP.
XX
XX AAX99259;
XX
XX 04-OCT-1999 (first entry)
XX
XX Nucleotide sequence encoding a peptide containing a site cleaved by OmpT.
XX
XX Recombinant peptide production: fusion peptide; accessory peptide;
KW chemical modification; agglutination; OmpT; ss.
XX
XX Synthetic.
XX
XX WO9338984-A1.
PN
PD 05-AUG-1999.
XX
XX 29-JAN-1999; 99WO-JP000406.
XX
XX 30-JAN-1998; 98JP-00032272.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohsuye K, Yabuta M, Suzuki Y;
PI
XX WPI; 1999-469333/39.
XX
XX Production of recombinant peptide as a fusion peptide followed by
PT

```

```

PT cleavage to give the target peptide in high yield and purity.
XX
XX Disclosure; Fig 6; 89pp; Japanese.
XX
XX The invention provides a method for the production of a desired
CC recombinant peptide where the recombinant peptide is produced as a fusion
CC peptide with a suitable accessory peptide. The fusion peptide is purified
CC from culture, subjected to any desired chemical modification, and then
CC cleaved and further purified to yield the desired peptide. The method is
CC used for efficient production of the desired peptides in high yield and
CC purity. The method allows the isoelectric point of the peptide purified
CC from the culture medium to be regulated to prevent agglutination of the
CC peptide and resultant loss of purity. Purity of above 98% and endotoxin
CC content below 0.03 U/mg can be achieved. The present sequence represents
CC a nucleotide sequence coding for an amino acid sequence containing a site
CC cleaved by OmpT
XX
XX Sequence 55 BP; 14 A; 16 C; 17 G; 8 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 0.0103 Length: 55
Score: 92.00 Matches: 15
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 88.24% Mismatches: 0
Query Match: 42.79% Indels: 0
DB: 2 Gaps: 0
US-10-500-018a-27 (1-36) x AAX99259 (1-55)
QY 13 HisGlyTyrAspAlaAspValArgLeuTyrArgArgHisHisGlySerGly 29
DB 53 CATGGTTATGACGGGAGCTCGCTGTATCGCTCATCACGGTTCGGA 3
RESULT 3
AAX99157/c
ID AAX99157 standard; DNA; 55 BP.
XX
XX AAX99157;
XX
XX 19-JAN-2001 (first entry)
XX
XX OmpT protease related oligonucleotide seugence SEQ ID NO:77.
XX
XX OmpT protease; cleavage; fusion protein; membrane protease; natriuretic;
KW Escherichia coli; PCR primer; ss.
XX
XX Escherichia coli.
OS Synthetic.
XX
XX WO200052193-A1.
PN
PD 08-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP001309.
XX
XX 04-MAR-1999; 99JP-00057731.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Okuno K, Yabuta M, Ohsuye K;
PI
XX WPI; 2000-579291/54.
XX
XX Controlled cleavage of peptides by OmpT protease by amino acid
PT substitution for ensuring cleavage only at desired site in fission of
XX fusion proteins.
XX
XX Example 1; Fig 2; 144pp; Japanese.
XX
XX The present invention describes a method for regulating the cleavage
CC sites of polypeptides by OmpT protease by preventing cleavage at unwanted
CC sites by converting the amino acid residue at position +1 to the site to
XX a specifically defined amino acid (where the residue at position -1 to
CC

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 09:09:52 ; Search time 1773 Seconds
(without alignments)
983.862 Million cell updates/sec

Title: US-10-500-018a-27
Perfect score: 215
Sequence: 1 BPHHHGPGQMGHYDADVRLYRRHHGSGSPSRHRR 36

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xl
-MODEL=frame_p2n.model -DEV=xl
-Q=/cgn2_1/USPTO_spool_p/US10500018/runat_25032005_092622_24160/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10500018 @CGN 1.1 5600 @runat_25032005_092622_24160 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.ev.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
C 1	88.5	41.2	300774	1 AE016912 Chromobac
2	84.5	39.3	14660	1 AF218037 Aeromonas
3	84	39.1	29850	1 AP005029 Streptomy
C 4	81.5	37.9	300704	1 AE017316 Desulfovi

5	37.7	2391	6	CQ591978	Sequence
6	37.7	2604	3	AY241926	Drosophil
C 7	37.7	6629	6	CQ591977	Sequence
8	37.7	41062	2	AC013168	Drosophil
C 9	37.7	164826	3	AC099038	Drosophil
10	37.7	168850	3	AC011070	Drosophil
C 11	37.7	300896	3	AE003499	Drosophil
12	37.7	300896	3	AF283839	Burkholder
13	37.7	300896	3	AF283839	Burkholder
14	37.7	300896	3	AF283839	Burkholder
15	37.7	300896	3	AF283839	Burkholder
16	37.7	300896	3	AF283839	Burkholder
17	37.7	300896	3	AF283839	Burkholder
18	37.7	300896	3	AF283839	Burkholder
19	37.7	300896	3	AF283839	Burkholder
20	37.7	300896	3	AF283839	Burkholder
21	37.7	300896	3	AF283839	Burkholder
22	37.7	300896	3	AF283839	Burkholder
23	37.7	300896	3	AF283839	Burkholder
24	37.7	300896	3	AF283839	Burkholder
25	37.7	300896	3	AF283839	Burkholder
26	37.7	300896	3	AF283839	Burkholder
27	37.7	300896	3	AF283839	Burkholder
28	37.7	300896	3	AF283839	Burkholder
29	37.7	300896	3	AF283839	Burkholder
30	37.7	300896	3	AF283839	Burkholder
31	37.7	300896	3	AF283839	Burkholder
32	37.7	300896	3	AF283839	Burkholder
33	37.7	300896	3	AF283839	Burkholder
34	37.7	300896	3	AF283839	Burkholder
35	37.7	300896	3	AF283839	Burkholder
36	37.7	300896	3	AF283839	Burkholder
37	37.7	300896	3	AF283839	Burkholder
38	37.7	300896	3	AF283839	Burkholder
39	37.7	300896	3	AF283839	Burkholder
40	37.7	300896	3	AF283839	Burkholder
41	37.7	300896	3	AF283839	Burkholder
42	37.7	300896	3	AF283839	Burkholder
43	37.7	300896	3	AF283839	Burkholder
44	37.7	300896	3	AF283839	Burkholder
45	37.7	300896	3	AF283839	Burkholder

ALIGNMENTS

RESULT 1	AE016912/c	300774 bp	DNA	linear	BCT 14-OCT-2003
LOCUS	Chromobacterium violaceum ATCC 12472	section 3 of 16	of the		
DEFINITION	complete genome.				
ACCESSION	AE016912				
VERSION	AE016912.1	GI:34101873			
KEYWORDS					
SOURCE	Chromobacterium violaceum ATCC 12472				
ORGANISM	Chromobacterium violaceum ATCC 12472				
REFERENCE	1 (bases 1 to 300774)				
AUTHORS	Brazilian National Genome Project Consortium				
CONSRMT	The complete genome sequence of Chromobacterium violaceum reveals				
TITLE	remarkable and exploitable bacterial adaptability				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)				
MEDLINE	22822880				
PUBMED	14500782				
REFERENCE	2 (bases 1 to 300774)				
AUTHORS	Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de				
CONSRMT	Almeida, L.G.P., de Almeida, R., Goncalves, J.A.A., Andrade, E.M.,				
TITLE	Antonio, R.V., Araripe, J., de Araujo, M.F.F., Filho, S.A., Azevedo, V.,				
JOURNAL	Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Borg, C.,				
MEDLINE	Blaney, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A.,				
PUBMED	Brocchi, M., Burity, H.A., Camargo, A., Cardoso, D.D.P.,				
REFERENCE	Carneiro, N.P., Carriaro, D.M., Carvalho, C.M.B., Cascado, J.C.M.,				

Cavada, B.S., Chueire, L.M.O., Pasa, T.B.C., Duran, N., Fagundes, N.,
 Falcao, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S.,
 Freitas, N.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,
 Freitas, N.S.A., Furian, L.R., Gazzinelli, R.T., Gomes, E.A.,
 Goncalves, P.R., Granger, I.B., Grattapaglia, D., Grissard, E.C.,
 Guimaraes, C.T., Hamann, E.S., Jungria, M., Jardim, S.N., Laurino, J.,
 Leoi, L.C.T., Passari, A.B., Lima, A., Loureiro, M.F., Lyra, M.C.P.,
 Macedo, M., Madeira, H.M.S., Manlio, G.P., Maranhao, A.Q.,
 Martins, W.S., Maturo, S.M.Z., de Medeiros, S.R.B., Meisner, R.V.,
 Menck, C.F.M., Moraes, M.M., Nascimento, F.F., Nicolas, M.F.,
 Oliveira, J.G., Oliveira, S.C., Paixao, R.F.C., Parente, J.A.,
 Pedrosa, F.O., Pena, S.P., Pereira, J.O., Pereira, M.,
 Pinto, L.S., Pinto, L.S., Porto, J.R., Potrich, D.P.,
 Neto, C.E.R., Reis, A.M.M., Rigor, R.U., Rondinelli, E., dos
 Santos, E.B.P., Santos, F.R., Schider, M.P.C., Seunares, H.N.,
 Silva, A.M.R., da Silva, A.L.C., Silva, D.W., Silva, R., Simoes, I.C.,
 Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.B.,
 Souza, R.C., Steffens, M.B.R., Steindler, M., Teixeira, S.R.,
 Urmenyi, T., Vettore, A., Wassem, R., Zana, A. and Simpson, A.J.G.
 Direct Submission
 Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de
 Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ
 25651070, Brazil
 FEATURES
 source
 1. 300774
 /organism="Chromobacterium violaceum ATCC 12472"
 /mol_type="genomic DNA"
 /strain="ATCC 12472"
 /db_xref="taxon:243365"
 complement (68. 598)
 /locus_tag="CV0566"
 complement (68. 598)
 /locus_tag="CV0566"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58241.1"
 /db_xref="GI:34101874"
 /translation="MLTDLTGKAGAPSPQPLEMLQACHKIRFCQDLQKLPYI
 EHGVDENARTIDVVRYPVAGPAHHTDEERELPLIDRVPVTASSLEQLSAHG
 YLHSCWNAIRDLDVALRDGRIKQISRNELOFARQYREHAAEAMLFPTASALSPE
 ELRQAGANMARRQDA"
 1267. 1812
 /gene="ampD"
 /locus_tag="CV0566"
 1267. 1812
 /gene="ampD"
 /locus_tag="CV0566"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/COG3023"
 /codon_start=1
 /transl_table=11
 /product="anhydro-N-acetylmuramyl-tripeptide amidase (ampD
 protein)"
 /protein_id="AAQ58242.2"
 /db_xref="GI:34330281"
 /translation="MTLGDGWGQARQLPSPNCDSRPGCAPELLVTHNISLPYRY
 GGAGVEQLFSNQLDPDEHPYKGIQOLRVSSHFFIRDSQLQLQFVPGKRAHGVSN
 WRGKCKNDISIGVEMEGCDPFPFSEAQYRTLAALSRELRALPLSAIAGHEHIAPCR
 KTDPCPWFWRRAQADSGLGF"
 complement (1817. 2272)
 /locus_tag="CV0567"
 complement (1817. 2272)
 /locus_tag="CV0567"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58243.1"
 /db_xref="GI:34101876"
 /translation="MSALTIEDIDADAITRLVNTAKTYLVPLDRVIRLQCLQGLPRD
 EILLAGVAGKARRRFPALLAQYDFSDVAVCGDFRQHVVFDRSVGFGVAGTRVDLP
 FRELLGNLDAISRLSDEAAFLGYSVALFALPGCELSFGDHPRPYKLK"
 complement (2269. 3597)

CDS
 /locus_tag="CV0568"
 complement (2269. 3597)
 /locus_tag="CV0568"
 /EC_number="4.1.3.27"
 /note="identified by sequence similarity; putative; ORF
 located using Glimmer/Blastx/COG0147"
 /codon_start=1
 /transl_table=11
 /product="probable anthranilate synthase"
 /protein_id="AAQ58244.1"
 /db_xref="GI:34101877"
 /translation="MNKIRLHVSVPADGRQTFLFALRCNYDYSYGKYLHERAVERAA
 LSLFFADRIALPLRSNGVRIETLAPLGLSPDPAAPGAFRLPDDPAPLAERLQO
 LDLPFADPADPRFLGYFFCYELLHAKLGNAGLGVFVHRLVDEGAVDCLIDP
 ADAGGSPHDLALGMLGADDFAAGRGDLYERYMYAIASRRIDAGLALSVNPSFAIQRCG
 ARDAYRHHQDLARNPAPYNLVFDGGDFQLAGMSFAMFLRLGRGLATSPICCTIARG
 RGEAEAOVATLLASDKDFELECVRADMQAKEACEDIRVDIEREVRANVPH
 SASVARKFKPTLIADICDHLWPATVGTPTDAAARLLAEHESRRWYAGAFYLCA
 EVELGTMTALLADGMATTVRGSTLSARSSPQLEQSELEAKASLILGLV"
 complement (3594. 4337)
 /locus_tag="CV0569"
 complement (3594. 4337)
 /locus_tag="CV0569"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58245.1"
 /db_xref="GI:34101878"
 /translation="MKSFKNLGRALAFIDPIVLVSLFLLVAAIASNLTSYKVIQVFG
 TEASLGFTFMPIILMLNPIAEVYKGRNQILFATLAEILFVAFSTALSQSDCG
 LLPLAERGHCEALNQSVLISEHIVKGSISPAIGCLIGSLNTRFLLYKLSWSSRLY
 FIRDFISSVGEVYTAICFMIAFYGVFPFATILKIFAFSLMFKFSTVILSWISQYI
 VLLIYRQAWSEGAPGGRKVFSSRYLQ"
 complement (4334. 4861)
 /locus_tag="CV0570"
 complement (4334. 4861)
 /locus_tag="CV0570"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58246.1"
 /db_xref="GI:34101879"
 /translation="MQAKNLKPHSVLHFEFKLNALRRIRDRKLSISELARITGINKQST
 LKRGLEFDRRLTSPNAHAIHGALINIGGETPMPAPVITGLDQDLALSEGROQPVWDE
 FILLEFSDLSALAVDQSLFRQLPFRRAVVVLTQTAHPRGRRLVYVDHGALSLEDSGR
 SAIGAVAGILFRKNA"
 4993. 5178
 /locus_tag="CV0571"
 4993. 5178
 /locus_tag="CV0571"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58247.1"
 /db_xref="GI:34101880"
 /translation="MIYVICGLVVLVVKLEINPDSLEGGFQGFGLVSHDYFCMSNLK
 LSGILKPLRVFLILCF"
 5297. 5863
 /locus_tag="CV0572"
 5297. 5863
 /locus_tag="CV0572"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAQ58248.1"
 /db_xref="GI:34101881"
 /translation="MESAFTVQTSRGVLDVQDPDDIDTADIAHHLAHLCRFGGACR
 QYVSEAHGVRMAETLPELKLGLLGMAYAGAVRGGCGQAPGQHLKTHAVGA
 RGLHSLGDRAELMADARLLATERRDLMPDENIEWFILTVGTQFLAERLWFMITAGR
 LASPCAVVQCFANSSSAQIRLQPSVR"
 complement (5930. 7795)
 /gene="kup2"
 /locus_tag="CV0573"

gene
 CDS
 /locus_tag="CV0573"

gene
 CDS
 /locus_tag="CV0573"

gene
 CDS
 /locus_tag="CV0573"

gene
 CDS
 /locus_tag="CV0573"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:42:02 ; Search time 94 Seconds
(without alignments)
126.805 Million cell updates/sec

Title: US-10-500-018a-27

Perfect score: 36

Sequence: 1 EPHHHPGGQMGYDADVRLYRRHHGSGPSRHR 36

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1407402 seqs, 331100923 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1407402

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	22.2	1661	US-10-437-963-155752	Sequence 155752,
2	7	19.4	18	US-10-105-717-3	Sequence 3, Appli
3	7	19.4	18	US-10-271-145-3	Sequence 3, Appli
4	7	19.4	18	US-10-100-699-3	Sequence 3, Appli
5	7	19.4	23	US-10-105-717-5	Sequence 5, Appli
6	7	19.4	23	US-10-105-717-5	Sequence 5, Appli
7	7	19.4	143	US-10-271-145-5	Sequence 174634,
8	7	19.4	162	US-10-437-963-174634	Sequence 226733,
9	7	19.4	175	US-10-424-599-226733	Sequence 41871, A
10	7	19.4	233	US-10-425-114-41871	Sequence 107544,
11	7	19.4	249	US-10-437-963-107544	Sequence 45053, A
12	7	19.4	306	US-10-425-114-45053	Sequence 7, Appli
13	7	19.4	306	US-10-630-636-7	Sequence 161967,
				US-10-437-963-161967	

14 7 19.4 371 15 US-10-404-667-2 Sequence 2, Appli
15 7 19.4 371 15 US-10-404-667-5 Sequence 5, Appli
16 7 19.4 391 14 US-10-057-531A-7 Sequence 7, Appli
17 7 19.4 391 14 US-10-057-532A-7 Sequence 7, Appli
18 7 19.4 393 14 US-10-057-531A-3 Sequence 3, Appli
19 7 19.4 393 14 US-10-057-532A-3 Sequence 3, Appli
20 7 19.4 434 15 US-10-425-114-68094 Sequence 68094, A
21 7 19.4 478 14 US-10-105-717-2 Sequence 2, Appli
22 7 19.4 478 14 US-10-105-717-2 Sequence 2, Appli
23 7 19.4 761 15 US-10-369-493-5391 Sequence 5391, Ap
24 7 19.4 844 16 US-10-437-963-135045 Sequence 135045,
25 7 19.4 1010 9 US-09-759-152-2 Sequence 2, Appli
26 7 19.4 1023 15 US-10-410-012-7 Sequence 7, Appli
27 7 19.4 1024 9 US-09-879-257A-30 Sequence 30, Appli
28 7 19.4 1024 15 US-10-369-493-724 Sequence 724, App
29 7 19.4 1024 15 US-10-460-524-9 Sequence 9, Appli
30 7 19.4 1032 16 US-10-416-708A-42 Sequence 42, Appli
31 7 19.4 1041 14 US-10-102-143-15 Sequence 15, Appli
32 7 19.4 1041 14 US-10-102-143-17 Sequence 17, Appli
33 7 19.4 1130 14 US-10-338-411-31 Sequence 31, Appli
34 7 19.4 1130 15 US-10-389-640-31 Sequence 31, Appli
35 7 19.4 1191 16 US-10-433-577-24 Sequence 24, Appli
36 7 19.4 1205 16 US-10-433-577-26 Sequence 26, Appli
37 6 16.7 12 15 US-10-441-908-24 Sequence 24, Appli
38 6 16.7 21 17 US-10-924-029-5 Sequence 5, Appli
39 6 16.7 29 10 US-09-962-756-2202 Sequence 2202, Ap
40 6 16.7 29 15 US-10-253-471-2202 Sequence 2202, Ap
41 6 16.7 29 15 US-10-253-493-2202 Sequence 2202, Ap
42 6 16.7 31 10 US-09-962-756-2203 Sequence 2203, Ap
43 6 16.7 31 15 US-10-253-471-2203 Sequence 2203, Ap
44 6 16.7 31 15 US-10-253-493-2203 Sequence 2203, Ap
45 6 16.7 32 9 US-09-864-761-35113 Sequence 35113, A

ALIGNMENTS

RESULT 1
US-10-437-963-155752
; Sequence 155752, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 155752
; LENGTH: 1661
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1661)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55487C.1.pap
US-10-437-963-155752

Query Match 22.2%; Score 8; DB 16; Length 1661;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 SGSPSRHR 35

Db 78 SGSPSRHR 85
|||||

RESULT 2

US-10-105-717-3
; Sequence 3, Application US/10105717
; Publication No. US2003032787A1
; GENERAL INFORMATION:
/4

; APPLICANT: Lanar, David E.
; APPLICANT: Dutta, Sheetij
; APPLICANT: Ware, Lisa A.
; APPLICANT: Nair, Lalitha P. V.
; TITLE OF INVENTION: Plasmodium falciparum AMA-1 Protein and Uses Thereof
; FILE REFERENCE: 003/246/SAP
; CURRENT APPLICATION NUMBER: US/10/105,717
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US 60/278,616
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: amino terminus 18 amino acids fused to AMA-1/E
US-10-105-717-3

Query Match 19.4%; Score 7; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHHHPGG 9
|||||

Db 5 HHHHPGG 11

RESULT 3

US-10-271-145-3
; Sequence 3, Application US/10271145
; Publication No. US20030138450A1
; GENERAL INFORMATION:
/4

; APPLICANT: Lanar, David E.
; APPLICANT: Dutta, Sheetij
; APPLICANT: Ware, Lisa A.
; APPLICANT: Nair, Lalitha P. V.
; TITLE OF INVENTION: Plasmodium falciparum AMA-1 Protein and Uses Thereof
; FILE REFERENCE: 003/255/SAP
; CURRENT APPLICATION NUMBER: US/10/271,145
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 10/105,717
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: amino terminus 18 amino acids fused to AMA-1/E
US-10-271-145-3

Query Match 19.4%; Score 7; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHHHPGG 9
|||||

Db 5 HHHHPGG 11

RESULT 4

US-10-100-699-3
; Sequence 3, Application US/10100699
; Publication No. US20030157650A1
; GENERAL INFORMATION:
/4
; APPLICANT: Lanar, David E.
; APPLICANT: Dutta, Sheetij
; APPLICANT: Ware, Lisa A.
; TITLE OF INVENTION: Recombinant P. vivax Merozoite Protein-1 p42 Vaccine
; FILE REFERENCE: 003/248/SAP
; CURRENT APPLICATION NUMBER: US/10/100,699
; CURRENT FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: US 60/277,002
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: vector sequence
US-10-100-699-3

Query Match 19.4%; Score 7; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHHHPGG 9
|||||

Db 5 HHHHPGG 11

RESULT 5

US-10-105-717-5
; Sequence 5, Application US/10105717
; Publication No. US20030032787A1
; GENERAL INFORMATION:
/4
; APPLICANT: Lanar, David E.
; APPLICANT: Dutta, Sheetij
; APPLICANT: Ware, Lisa A.
; APPLICANT: Nair, Lalitha P. V.
; TITLE OF INVENTION: Plasmodium falciparum AMA-1 Protein and Uses Thereof
; FILE REFERENCE: 003/246/SAP
; CURRENT APPLICATION NUMBER: US/10/105,717
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US 60/278,616
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 5
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: N-terminal 23 amino acids in recombinant AMA-1/E
US-10-105-717-5

Query Match 19.4%; Score 7; DB 14; Length 23;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHHHPGG 9
|||||

Db 4 HHHHPGG 10

RESULT 6

US-10-271-145-5
; Sequence 5, Application US/10271145
; Publication No. US20030138450A1

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:36:36 ; Search time 44 seconds
(without alignments)
61.077 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 36

Sequence: 1 EPHHHPGGQMHGYDADVRLYRRHHGSGSPSRHRR 36

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Issued Patents_AA:*
- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	25.0	15	2	US-08-811-028-52
2	8	22.2	404	4	US-09-252-991A-33014
3	7	19.4	12	3	US-08-523-373-16
4	7	19.4	15	1	US-08-352-179-22
5	7	19.4	16	3	US-08-523-373-17
6	7	19.4	17	3	US-08-523-373-18
7	7	19.4	18	3	US-08-523-373-19
8	7	19.4	20	3	US-08-523-373-1
9	7	19.4	20	3	US-08-523-373-20
10	7	19.4	22	3	US-08-523-373-21
11	7	19.4	24	1	US-08-657-192-10
12	7	19.4	26	1	US-08-352-179-21
13	7	19.4	30	1	US-08-352-179-20
14	7	19.4	37	1	US-08-352-179-19
15	7	19.4	139	2	US-08-805-918-2
16	7	19.4	163	5	PCT-US93-08435-2
17	7	19.4	248	4	US-09-252-991A-30679
18	7	19.4	344	1	US-08-657-192-3
19	7	19.4	344	3	US-08-523-373-5
20	7	19.4	392	3	US-08-523-373-6
21	7	19.4	465	4	US-09-252-991A-24848
22	7	19.4	516	4	US-09-489-039A-10117
23	7	19.4	532	1	US-08-657-192-9
24	7	19.4	532	3	US-08-523-373-7
25	7	19.4	537	1	US-08-657-192-15
26	7	19.4	1010	4	US-09-654-449-2
27	7	19.4	1010	4	US-09-759-152A-2

RESULT 1

US-08-811-028-52
; Sequence 52, Application US/08811028C
; Patent No. 5891671
; GENERAL INFORMATION:
; APPLICANT: SUZUKI, Yuji
; APPLICANT: MAGOTA, Koji
; APPLICANT: MASUDA, Toyofumi
; TITLE OF INVENTION: METHOD FOR CLEAVING CHIMERIC ENZYME USING PROCESSING
; TITLE OF INVENTION: ENZYME
; FILE REFERENCE: 001560-294
; CURRENT APPLICATION NUMBER: US/08/811,028C
; CURRENT FILING DATE: 1987-03-04
; EARLIER APPLICATION NUMBER: JP 8-70906
; EARLIER FILING DATE: 1996-03-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:modified
; OTHER INFORMATION: Kex2-660 recognition site; x is gly, ala, val,
; OTHER INFORMATION: leu, ile, ser, thr, asp, glu, asn, gin, lys, arg,
; OTHER INFORMATION: phe, tyr, trp, his, or pro
US-08-811-028-52

Query Match 25.0%; Score 9; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PHHHPGGR 10

Db 1 PHHHPGGR 9

RESULT 2

US-09-252-991A-33014
; Sequence 33014, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 33014
 ; LENGTH: 404
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-33014

Query Match 22.2%; Score 8; DB 4; Length 404;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 23 RRRHSGS 30
 Db 224 RRRHSGS 231

RESULT 3
 US-08-523-373-16
 ; Sequence 16, Application US/08523373
 ; Patent No. 6037145
 ; GENERAL INFORMATION:
 ; APPLICANT: Yabuta, Masayuki
 ; APPLICANT: OHSUYE, Kazuhiro
 ; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
 ; STREET: 699 Prince Street
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22314-3187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/523,373
 ; FILING DATE: 05-SEP-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 6-238595
 ; FILING DATE: 07-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 6-296028
 ; FILING DATE: 07-NOV-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meuth, Donna M.
 ; REGISTRATION NUMBER: 36,607
 ; REFERENCE/DOCKET NUMBER: 001560-251
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-836-6620
 ; TELEFAX: 703-836-2021
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 12 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: C-terminal
 US-08-523-373-16

Query Match 19.4%; Score 7; DB 3; Length 12;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 RLYRRHH 26
 Db 5 RLYRRHH 11

RESULT 4
 US-08-352-179-22
 ; Sequence 22, Application US/08352179
 ; Patent No. 5670340
 ; GENERAL INFORMATION:
 ; APPLICANT: YABUTA, Masayuki
 ; APPLICANT: SUZUKI, Yuji
 ; APPLICANT: OHSUYE, Kazuhiro
 ; APPLICANT: OSHIMA, Takehiro
 ; APPLICANT: ONAI, Seiko
 ; APPLICANT: MAGOTA, Koji
 ; APPLICANT: TANAKA, Shoji
 ; TITLE OF INVENTION: PROCESS FOR PRODUCING PEPTIDE
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Burns, Doane, Swecker and Mathis
 ; STREET: The George Mason Bldg., Washington & Prince
 ; STREET: Sts.
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA: US/08/352,179
 ; APPLICATION NUMBER: US/08/352,179
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/929,597
 ; FILING DATE: 17-AUG-1992
 ; APPLICATION NUMBER: JP 3-320769
 ; FILING DATE: 19-AUG-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 4-223520
 ; FILING DATE: 31-JUL-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Crane-Feury, Sharon E
 ; REGISTRATION NUMBER: 36,113
 ; REFERENCE/DOCKET NUMBER: 001560-175
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 836-6620
 ; TELEFAX: (703) 836-2021
 ; INFORMATION FOR SEQ ID NO: 22:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-352-179-22

Query Match 19.4%; Score 7; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 RLYRRHH 26
 Db 3 RLYRRHH 9

RESULT 5
 US-08-523-373-17
 ; Sequence 17, Application US/08523373
 ; Patent No. 6037145
 ; GENERAL INFORMATION:
 ; APPLICANT: Yabuta, Masayuki
 ; APPLICANT: Ohsuye, Kazuhiro

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:28:46 ; Search time 118 Seconds
(without alignments)
156.228 Million cell updates/sec

Title: US-10-500-018a-27
Perfect score: 36
Sequence: 1 EPHHHHGGQMHGYDADVRLYRRHHGSGSPSRHR 36

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	19.4	147	2 Q47340	Q47340 escherichia
2	7	19.4	186	2 Q680T8	Q680T8 arabidopsis
3	7	19.4	187	2 Q8L921	Q8L921 arabidopsis
4	7	19.4	187	2 Q9LTS1	Q9LTS1 arabidopsis
5	7	19.4	260	2 Q86RL9	Q86RL9 ilyanassa o
6	7	19.4	306	2 Q8LH59	Q8LH59 oryza sativ
7	7	19.4	385	2 Q7QE88	Q7QE88 anopheles g
8	7	19.4	390	2 Q84SU8	Q84SU8 oryza sativ
9	7	19.4	619	2 Q979T6	Q979T6 thermoplasm
10	7	19.4	679	2 Q6PUC7	Q6PUC7 anopheles g
11	7	19.4	825	2 Q7Q236	Q7Q236 anopheles g
12	7	19.4	907	2 Q6C9S1	Q6C9S1 yarrowia li
13	7	19.4	1023	1 BGAL_ECOLI	P00722 escherichia
14	7	19.4	1024	2 Q8FKG6	Q8FKG6 escherichia
15	7	19.4	1024	2 Q8X685	Q8X685 escherichia
16	7	19.4	1029	2 Q8VNN2	Q8VNN2 escherichia
17	7	19.4	1202	1 LPAL_HUMAN	Q13136 homo sapien
18	7	19.4	1226	1 CDN1_HUMAN	Q81WY9 homo sapien
19	7	19.4	1227	2 Q6NYD0	Q6NYD0 homo sapien
20	7	19.4	7576	2 Q9ZGA4	Q9ZGA4 streptomyce
21	6	15.7	66	2 Q8BMX9	Q8BMX9 mus musculu
22	6	16.7	75	2 Q6ZBC7	Q6ZBC7 oryza sativ
23	6	16.7	83	2 Q914R1	Q914R1 pseudomonas
24	6	16.7	94	2 Q08302	Q08302 rattus norv
25	6	16.7	99	2 Q8FP19	Q8FP19 corynebacte
26	6	16.7	101	2 Q8NL05	Q8NL05 xanthomonas
27	6	16.7	101	2 Q7CLJ4	Q7CLJ4 xanthomonas
28	6	16.7	113	2 Q83DE2	Q83DE2 coxiella bu
29	6	16.7	114	2 Q9YC18	Q9YC18 aeropyrum p
30	6	16.7	118	2 Q01614	Q01614 pneumocysti
31	6	16.7	119	2 Q6CGD6	Q6CGD6 yarrowia li

32	6	16.7	119	2 Q8GY07	Q8GY07 arabidopsis
33	6	16.7	119	2 Q9SJP7	Q9SJP7 arabidopsis
34	6	16.7	123	2 Q6IDP7	Q6IDP7 arabidopsis
35	6	16.7	124	2 Q9LGR2	Q9LGR2 oryza sativ
36	6	16.7	125	1 HNT1_HUMAN	P49773 homo sapien
37	6	16.7	131	2 Q9LUK5	Q9LUK5 arabidopsis
38	6	16.7	136	2 Q8A649	Q8A649 bacteroides
39	6	16.7	139	2 Q8S6Y8	Q8S6Y8 oryza sativ
40	6	16.7	146	2 Q7R005	Q7R005 giardia lam
41	6	16.7	154	2 Q7MSX3	Q7MSX3 mus musculu
42	6	16.7	155	1 RL24_ASHGO	Q752u6 ashbya goss
43	6	16.7	155	1 RL24_CANGA	Q6fxy9 candida gla
44	6	16.7	155	1 RL24_KIULA	P38665 kluyveromyc
45	6	16.7	158	2 Q7X8Q6	Q7X8Q6 oryza sativ

ALIGNMENTS

RESULT 1

Q47340 PRELIMINARY; PRT; 147 AA.
AC Q47340
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE LacZ 5'-region (fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89362462; PubMed=2475637;
RA Ruteshouser E.C., Richardson J.P.;
RT "Identification and characterization of transcription termination
RT sites in the Escherichia coli lacZ gene.";
RL J. Mol. Biol. 208:23-43(1989).
DR EMBL; X16313; CAA34380.1; -.
DR HSSP; P00722; 1BGU.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006104; Glyco hydro 2SB.
DR Pfam; PF02837; Glyco_hydro_2_N; I.
FT NMAN TER 147 147
SQ SEQUENCE 147 AA; 16796 MW; E801CCL57CE07E72 CRC64;

Query Match 19.4%; Score 7; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 QMHGYDA 17
Db |||||
92 QMHGYDA 98

RESULT 2

Q680T8 PRELIMINARY; PRT; 186 AA.
AC Q680T8
DT 25-OCT-2004 (Tremblrel. 28, Created)
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE Hypothetical protein At5g57000.
GN Name=At5g57000;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

RA Totoki Y., Seki M., Ishida J., Nakajima M., Enju A., Kamiya A.,
 RA Narusaka M., Shin-i T., Nakagawa M., Sakamoto N., Oishi K., Kohara Y.,
 RA Kobayashi M., Toyoda A., Sakaki Y., Sakurai T., Iida K., Akiyama K.,
 RA Satou M., Toyoda T., Konagaya A., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.,
 RT "Large-scale analysis of RIKEN Arabidopsis full-length (RAFT) cDNAs.",
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AK175779; BAD43542.1; --
 KW Hypothetical protein
 SQ SEQUENCE 186 AA; 20898 MW; 7D895113A67E7754 CRC64;

Query Match 19.4%; Score 7; DB 2; Length 186;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 GSGSPSR 33
 Db 146 GSGSPSR 152

RESULT 3

Q8L921 ID Q8L921 PRELIMINARY; PRT; 187 AA.
 AC Q8L921;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hypothetical protein
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22088475; PubMed=12093376;
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 3:RESEARCH0029-RESEARCH0029 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AY088676; AAM66998.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 187 AA; 21022 MW; 7BD5D3ADD63534FC CRC64;

Query Match 19.4%; Score 7; DB 2; Length 187;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 GSGSPSR 33
 Db 147 GSGSPSR 153

RESULT 4

Q9LTS1 ID Q9LTS1 PRELIMINARY; PRT; 187 AA.
 AC Q9LTS1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE Gb|AAF21159.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:31-63 (2000).
 DR EMBL; AB024035; BAA97030.1; --
 SQ SEQUENCE 187 AA; 21036 MW; 7BD4621DD63535FD CRC64;

Query Match 19.4%; Score 7; DB 2; Length 187;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 GSGSPSR 33
 Db 147 GSGSPSR 153

RESULT 5

Q86RL9 ID Q86RL9 PRELIMINARY; PRT; 260 AA.
 AC Q86RL9;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Even-skipped transcription factor-like protein (fragment).
 GN Name-Eve;
 OS Ilyanassa obsolleta.
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
 OC Apogastropoda; Caenogastropoda; Sorbeoconcha; Hypsogastropoda;
 OC Neogastropoda; Succinoidea; Nasseariidae; Ilyanassa.
 OX NCBI_TaxID=34582;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22366420; PubMed=12478296; DOI=10.1038/nature01241;
 RA Lambert J.D., Nagy L.M.;
 RT "Asymmetric inheritance of centrosomally localized mRNAs during
 RT embryonic cleavages.";
 RL Nature 420:682-686 (2002).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AF499912; AAC20893.1; --
 DR HSSP; P06602; 1JGG
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_lambdomain_like.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 260 260
 SQ SEQUENCE 260 AA; 27586 MW; F1F65FC0AF77483D CRC64;

Query Match 19.4%; Score 7; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PHHHHPG 8
 Db 230 PHHHHPG 236

RESULT 6

Q8LH59 ID Q8LH59 PRELIMINARY; PRT; 306 AA.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:34:51 ; Search time 40 Seconds
(without alignments)
86.595 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 36

Sequence: 1 EPHHHHPGQRMHGVDVRLYRRHHGSGPSRRHR 36

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	19.4	761	2 T15776	hypothetical prote
2	7	19.4	1024	1 GBEC	beta-galactosidase
3	7	19.4	1024	2 E90678	beta-D-galactosida
4	7	19.4	1024	2 A85529	beta-D-galactosida
5	7	19.4	1202	2 S55553	LAR-interacting pr
6	7	19.4	7576	2 T17428	FK506 polyketide s
7	6	16.7	83	2 E83512	hypothetical prote
8	6	16.7	94	2 S33811	cp91 protein - rat
9	6	16.7	114	2 E72600	hypothetical prote
10	6	16.7	172	2 F95895	hypothetical prote
11	6	16.7	194	2 T52044	dof zinc finger pr
12	6	16.7	198	2 A96811	unknown protein T1
13	6	16.7	210	2 A37358	oncofetal protein
14	6	16.7	239	2 S50891	hypothetical prote
15	6	16.7	243	2 F96595	unknown protein, 2
16	6	16.7	255	2 T48543	hypothetical prote
17	6	16.7	267	2 A92748	hypothetical prote
18	6	16.7	268	2 F97529	hypothetical prote
19	6	16.7	271	2 A95267	probable ABC trans
20	6	16.7	273	2 PC1273	30K phycocyanin ro
21	6	16.7	283	2 T36121	probable secreted
22	6	16.7	285	2 T18689	hypothetical prote
23	6	16.7	289	2 S30257	phycocyanin linker
24	6	16.7	291	2 S33209	extensin-like prot
25	6	16.7	295	2 A87703	citrate lyase beta
26	6	16.7	297	2 H86519	hypothetical prote
27	6	16.7	297	2 G81565	hypothetical prote
28	6	16.7	297	2 E72102	5'-methylthioadeno
29	6	16.7	320	2 S48367	hypothetical prote

30	6	16.7	357	2 G70010	probable aspartate
31	6	16.7	359	2 B86409	p3H9.6 protein - A
32	6	16.7	366	2 S53073	hypothetical prote
33	6	16.7	378	2 AD1541	hypothetical prote
34	6	16.7	391	2 E95947	probable PAD-depen
35	6	16.7	416	2 S52078	proatacycilin - rat
36	6	16.7	417	2 A54416	proatacycilin recep
37	6	16.7	422	2 JCS325	methicillin resist
38	6	16.7	425	2 JC7230	vitamin D receptor
39	6	16.7	427	2 T35066	tryptophan synthas
40	6	16.7	427	2 JN0785	Carbon catabolite
41	6	16.7	430	2 T04678	hypothetical prote
42	6	16.7	447	2 I56511	zic protein - mous
43	6	16.7	448	2 S17370	DNA-binding protei
44	6	16.7	467	2 F84699	probable flavonol
45	6	16.7	468	2 JN0016	peripherin interme

ALIGNMENTS

RESULT 1

T15776

hypothetical protein C34F11.9a - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C;Accession: T15776

R;Bentley, D.

submitted to the EMBL Data Library, January 1996

A;Description: The sequence of C. elegans cosmid C34F11.

A;Reference number: Z18400

A;Accession: T15776

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-761 <BEN>

A;Cross-references: EMBL:U46753; NID:g1166623; PID:g1166632; PIDN:AAA85766.1; CESP:C34F11

C;Genetics:

A;Gene: CESP:C34F11.9a

A;Introns: 41/1; 73/3; 123/2; 196/1; 228/1; 413/1; 438/2; 479/1; 559/2; 618/1; 693/1

Query Match 19.4%; Score 7; DB 2; Length 761;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 SPSRHR 36

Db 410 SPSRHR 416

RESULT 2

GBEC

beta-galactosidase (EC 3.2.1.23) lacZ [validated] - Escherichia coli (strain K-12)

N;Alternate names: beta-D-galactosidase; lactase

C;Species: Escherichia coli

C;Date: 24-Apr-1984 #sequence_revision 23-Sep-1997 #text_change 09-Jul-2004

C;Accession: A90981; A92233; A93224; S06878; T41218; H64761; I40987; A00898; S14637; S144

R;Kallings, A.; Otto, K.; Ruther, U.; Muller-Hill, B.

EMBO J. 2, 593-597, 1983

A;Title: Sequence of the lacZ gene of Escherichia coli.

A;Reference number: A90981; MUID:84028567; PMID:6313347

A;Accession: A90981

A;Molecule type: DNA

A;Residues: 2-1024 <KAL>

A;Cross-references: UNIPROT:P00722; GB:V00296; NID:g41901; PIDN:CAA23573.1; PID:g1197203

A;Note: translation of initiator Met is not shown

R;Fowler, A.V.; Zabin, I.

J. Biol. Chem. 253, 5521-5525, 1978

A;Title: Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and

A;Reference number: A92233; MUID:78218239; PMID:97298

A;Accession: A92233

A;Molecule type: protein

A;Residues: 2-1024 <FOW>

A;Note: this is the final paper in a series

R;Calos, M.P.; Miller, J.H.
 Nature 285, 38-41, 1980
 A;Title: Molecular consequences of deletion formation mediated by the transposon Tn9.
 A;Reference number: A93224; PMID:80188189; PMID:6246435
 A;Accession: A93224
 A;Molecule type: DNA
 A;Residues: 356-476 <CAL>
 R;Rutshouser, E.C.; Richardson, J.P.
 J. Mol. Biol. 208, 23-43, 1989
 A;Title: Identification and characterization of transcription termination sites in the E. coli trp operon.
 A;Reference number: S06878; PMID:89362462; PMID:2475637
 A;Accession: S06878
 A;Status: translation not shown
 A;Molecule type: mRNA
 A;Residues: 1-147 <RUT>
 A;Cross-references: EMBL:X16313; NID:g41903; PIDN:CAA34380.1; PID:g41904
 R;Mikryukov, N.N.; Petrov, N.A.; Karginov, V.A.; Vassilenko, S.K.
 Bioorg. Khim. 6, 1735-1736, 1980
 A;Title: Nucleotide sequence of a lambda-plac 5-1 DNA region coding for a COOH-terminal protein.
 A;Reference number: I41218
 A;Accession: I41218
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 967-971, 'R', 973-1022, 'E', 1024 <MIK>
 A;Cross-references: GB:M38327; NID:g146061; PIDN:AAA23835.1; PID:g146062
 R;Jacobson, R.H.; Zhang, X.; Dubose, R.F.; Matthews, B.W.
 submitted to the Brookhaven Protein Data Bank, July 1994
 A;Reference number: A65162; PDB:1BGL
 A;Contents: annotation; X-ray crystallography, 2.50 angstroms, residues 4-1024
 R;Jacobson, R.H.; Zhang, X.J.; DuBoise, R.F.; Matthews, B.W.
 Nature 369, 761-766, 1994
 A;Title: Three-dimensional structure of beta-galactosidase from E. coli.
 A;Reference number: A58594; PMID:94277211; PMID:8009071
 A;Contents: annotation; X-ray crystallography, 2.50 angstroms
 R;Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of Escherichia coli K-12.
 A;Reference number: A64720; PMID:97426617; PMID:9278503
 A;Accession: H64761
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-1024 <BLAT>
 A;Cross-references: GB:AE000141; GB:U00096; NID:g1786532; PIDN:AACT3447.1; PID:g1786539
 A;Experimental source: strain K-12, substrain MGL1655
 R;Prentki, P.
 Gene 122, 231-232, 1992
 A;Title: Nucleotide sequence of the classical lacZ deletion delta M15.
 A;Reference number: I40987; PMID:93083990; PMID:1339377
 A;Accession: I40987
 A;Molecule type: DNA
 A;Residues: 1-11, 43-50 <RES>
 A;Cross-references: EMBL:X58252; NID:g40882; PIDN:CAA41206.1; PID:g40883
 C;Genetics:
 A;Gene: lacZ
 A;Map position: 8 min
 C;Complex: homotetramer
 C;Function:
 A;Description: catalyzes hydrolysis of lactose into galactose and glucose
 C;Superfamily: beta-galactosidase
 A;Keywords: glycosidase; homotetramer; hydrolase; magnesium
 F;2-1024/Product: beta-galactosidase #status experimental <MAT>
 F;2-50/Region: alpha complementation
 F;51-218/Domain: 1, jelly-roll beta-barrel #status predicted <DM1>
 F;219-334/Domain: 2, fibronectin type-III fold #status predicted <DM2>
 F;335-627/Domain: 3, distorted TIM barrel #status predicted <DM3>
 F;628-737/Domain: 4, fibronectin type-III fold #status predicted <DM4>
 F;738-1024/Domain: 5, anti-parallel beta-sandwich #status predicted <DM5>
 F;417, 419, 462/Binding site: magnesium (Glu, His, Glu) #status experimental
 F;462, 504, 538/Active site: Glu, Tyr, Glu #status predicted

Query Match 19.4%; Score 7; DB 1; Length 1024;
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 QMHGYDA 17
 |||||
 Db 92 QMHGYDA 98
 RESULT 3
 E90678
 beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995)
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C;Accession: E90678
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Sasawara, N.T.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A;Reference number: A99629; PMID:21156231; PMID:11258796
 A;Accession: E90678
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1024 <HAY>
 A;Cross-references: UNIPROT:O8X685; GB:BA000007; PIDN:BA033820.1; PID:g13359854; GSPDB:GN
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs0397
 C;Superfamily: beta-galactosidase
 Query Match 19.4%; Score 7; DB 2; Length 1024;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 QMHGYDA 17
 |||||
 Db 92 QMHGYDA 98
 RESULT 4
 A85529
 beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C;Accession: A85529
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; PMID:21074935; PMID:11206551
 A;Accession: A85529
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1024 <STO>
 A;Cross-references: UNIPROT:O8X685; GB:AE000174; NID:g12513175; PIDN:AAG54693.1; GSPDB:GN
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: lacZ
 C;Superfamily: beta-galactosidase
 Query Match 19.4%; Score 7; DB 2; Length 1024;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 QMHGYDA 17
 |||||
 Db 92 QMHGYDA 98
 RESULT 5
 S5553
 LAR-interacting protein Lipib - human
 C;Species: Homo sapiens (man)
 C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
 C;Accession: S5553; S55552

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:27:51 ; Search time 121 Seconds
(without alignments)
115.069 Million cell updates/sec

Title: US-10-500-018a-27

Perfect score: 36

Sequence: 1 EPHHHGPGQMGYDADVRLYRRHHGSGSPSRHRR 36

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	36	8	Adl66848 Ghrelin-r
2	36	100.0	57	8	Adl66814 Human Ghr
3	34	94.4	36	8	Adl66847 Ghrelin-r
4	34	94.4	57	8	Adl66815 Human Ghr
5	34	94.4	57	8	Adl66813 Human Ghr
6	32	88.9	57	8	Adl66820 Human Ghr
7	32	88.9	57	8	Adl66821 Human Ghr
8	32	88.9	57	8	Adl66819 Human Ghr
9	30	83.3	57	8	Adl66816 Human Ghr
10	29	80.6	57	8	Adl66817 Human Ghr
11	18	50.0	69	8	Adl66818 Human Ghr
12	17	47.2	184	3	Aab23948 Plasmid p
13	15	41.7	26	3	Aab23986 Peptide a
14	15	41.7	154	2	Aay28959 Amino aci
15	15	41.7	184	2	Aay28961 Amino aci
16	15	41.7	184	2	Aay28962 Amino aci
17	13	36.1	13	8	Adl66851 Ghrelin a
18	12	33.3	14	8	Adl66856 Ghrelin a
19	12	33.3	26	3	Aab23988 Linker pe
20	11	30.6	13	8	Adl66853 Ghrelin a
21	11	30.6	13	8	Adl66852 Ghrelin a
22	11	30.6	20	2	Aay28957 Amino aci
23	11	30.6	20	3	Aab23984 Peptide a
24	10	27.8	187	2	Aay28960 Amino aci
25	9	25.0	9	3	Aab23967 OmPT prot

ALIGNMENTS

RESULT 1

ADL66848

ID ADL66848 standard; peptide; 36 AA.

AC ADL66848;

DT 20-MAY-2004 (first entry)

DE Ghrelin-related linker peptide used in fusion protein pl17 8-280RR.

XX drug; Ghrelin; growth hormone secretagogue; pl17 8-280RR; linker.

XX Unidentified.

XX WO2003084983-A1.

XX 16-OCT-2003.

XX 10-APR-2003; 2003WO-JP004590.

XX 11-APR-2002; 2002JP-00109761.

XX (DAI-) DAIICHI SUNTORY PHARMA CO LTD.

XX (KANG/) KANGAWA K.

XX Minamitake Y, Matsumoto M, Makino T;

XX WPI; 2004-098839/10.

XX Producing modified peptides or proteins with physiological activity comprises fusing side chain-modified peptide fragments obtained by solid-phase synthesis and non-modified peptides by genetic modification.

XX Claim 15; SEQ ID NO 27; 120pp; Japanese.

XX The invention relates to a novel method for producing protected or modified peptide fragments containing desired sequences of amino acids and/or non-amino acids, at least one of which is suitably protected for preparing the peptide fragment with use of a weakly acidic-eliminating resin to enable cleavage from the resin under weakly acidic conditions. The method of the invention may be used for producing modified peptide fragments or proteins and producing protected peptide fragments that do not contain any modified (non-)amino acids for application as drugs. The current sequence is that of a Ghrelin and growth hormone secretagogue-related linker peptide of the invention which was used in fusion protein pl17 8-280RR.

XX Sequence 36 AA;

SQ

Query Match 100.0%; Score 36; DB 8; Length 36;
Best Local Similarity 100.0%; Pred. No. 6.7e-26;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRHR 36
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRHR 36

RESULT 2
ADL66814
ID ADL66814 standard; protein; 57 AA.
AC ADL66814;
XX 20-MAY-2004 (first entry)
DT Human Ghrelin 8-28/ linker fusion protein p117s 8-28ORR.
DE drug; Ghrelin; growth hormone secretagogue; p117s 8-28ORR fusion; human;
KW linker.
XX Homo sapiens.
OS Unidentified.
OS Synthetic.
XX WO2003084983-A1.
XX 16-OCT-2003.
XX 10-APR-2003; 2003WO-JP004590.
XX 11-APR-2002; 2002JP-00109761.
XX (DAII-) DAIICHI SUNTORY PHARMA CO LTD.
PA (KANG/) KANGAWA K.
XX Minamitake Y, Matsumoto M, Makino T;
XX WPI; 2004-098839/10.
XX Producing modified peptides or proteins with physiological activity
PT comprises fusing side chain-modified peptide fragments obtained by solid-
phase synthesis and non-modified peptides by genetic modification.
XX Example 13; Fig 5; 120pp; Japanese.
XX The invention relates to a novel method for producing protected or
CC modified peptide fragments containing desired sequences of amino acids
CC and/or non-amino acids, at least one of which is suitably protected for
CC preparing the peptide fragment with use of a weakly acidic-eliminating
CC resin to enable cleavage from the resin under weakly acidic conditions.
CC The method of the invention may be used for producing modified peptide
CC fragments or proteins and producing protected peptide fragments that do
CC not contain any modified (non-)amino acids for application as drugs. The
CC current sequence is that of a human Ghrelin 8-28/ linker fusion protein
CC of the invention.
XX Sequence 57 AA;

Query Match 100.0%; Score 36; DB 8; Length 57;
Best Local Similarity 100.0%; Pred. No. 9.8e-26;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRHR 36
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRHR 36

RESULT 3
ADL66847
ID ADL66847 standard; peptide; 36 AA.

XX ADL66847;
XX 20-MAY-2004 (first entry)
XX Ghrelin-related linker peptide used in fusion protein p117 8-28oPR.
DE drug; Ghrelin; growth hormone secretagogue; p117 8-28oPR; linker.
XX Unidentified.
XX WO2003084983-A1.
XX 16-OCT-2003.
XX 10-APR-2003; 2003WO-JP004590.
XX 11-APR-2002; 2002JP-00109761.
XX (DAII-) DAIICHI SUNTORY PHARMA CO LTD.
PA (KANG/) KANGAWA K.
XX Minamitake Y, Matsumoto M, Makino T;
XX WPI; 2004-098839/10.
XX Producing modified peptides or proteins with physiological activity
PT comprises fusing side chain-modified peptide fragments obtained by solid-
phase synthesis and non-modified peptides by genetic modification.
XX Example 1; SEQ ID NO 26; 120pp; Japanese.
XX The invention relates to a novel method for producing protected or
CC modified peptide fragments containing desired sequences of amino acids
CC and/or non-amino acids, at least one of which is suitably protected for
CC preparing the peptide fragment with use of a weakly acidic-eliminating
CC resin to enable cleavage from the resin under weakly acidic conditions.
CC The method of the invention may be used for producing modified peptide
CC fragments or proteins and producing protected peptide fragments that do
CC not contain any modified (non-)amino acids for application as drugs. The
CC current sequence is that of a Ghrelin and growth hormone secretagogue-
CC related linker peptide of the invention which was used in fusion protein
CC p117 8-28oPR.
XX Sequence 36 AA;

Query Match 94.4%; Score 34; DB 8; Length 36;
Best Local Similarity 100.0%; Pred. No. 4.6e-24;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRH 34
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 EPHHHHPGGRQMGYDADVRLYRRHHGSGSPSRH 34

RESULT 4
ADL66815
ID ADL66815 standard; protein; 57 AA.

XX ADL66815;
XX 20-MAY-2004 (first entry)
XX Human Ghrelin 8-28/ linker fusion protein p117s 8-28oKR.
DE drug; Ghrelin; growth hormone secretagogue; p117s 8-28oKR fusion; human;
KW linker.
XX Homo sapiens.
OS Unidentified.
OS Synthetic.
XX WO2003084983-A1.

US-10-437-963-157386

```

; Sequence 157386, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 157386
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_56963C.1.pap
; US-10-437-963-157386

Query Match      32.3%; Score 69.5; DB 16; Length 932;
Best Local Similarity 38.6%; Pred. No. 4.2;
Matches 17; Conservative 1; Mismatches 15; Indels 11; Gaps 2;

QY 3 HHHHPGQRMHGYDADVRLYRRHHGSGPSR-----HRR 36
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 254 HHHHPGQRHQRG-DSDPRCEADRRDRDGRSGRGLSNLYGHRR 296
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 3
US-10-074-225A-5
; Sequence 5, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; FILE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 085596CD1
; US-10-074-225A-5

Query Match      30.7%; Score 66; DB 14; Length 148;
Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 14; Conservative 4; Mismatches 17; Indels 14; Gaps 1;

QY 2 PHHHHPGQRMHGYDADVRLYRRH-----HSGSGPSRHR 36
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 42 PHGHHPHGHHPHGHHPCHDFQDYGCDPPPHNQGHCHCGHGGPPPGHLR 90
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 4
US-10-437-963-111601
; Sequence 111601, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 111601
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15565C.1.pap
; US-10-437-963-111601

Query Match      30.7%; Score 66; DB 16; Length 183;
Best Local Similarity 32.1%; Pred. No. 2.2;
Matches 18; Conservative 5; Mismatches 11; Indels 22; Gaps 4;

QY 2 PHHHHP-----GGR-----OMHGY-DADVRLYR--RHGSGPSRHR 35
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 114 PHHHHPHYSVQHGRMSGGKAVSSYNYHVHGYFSDLDLHGYSGHHGGDVVPAAR 169
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 5
US-09-919-039-62
; Sequence 62, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 62
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 085596CD1
; US-09-919-039-62

Query Match      30.7%; Score 66; DB 10; Length 525;
Best Local Similarity 28.6%; Pred. No. 6.4;
Matches 14; Conservative 4; Mismatches 17; Indels 14; Gaps 1;

QY 2 PHHHHPGQRMHGYDADVRLYRRH-----HSGSGPSRHR 36
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 391 PHGHHPHGHHPHGHHPCHDFQDYGCDPPPHNQGHCHCGHGGPPPGHLR 439
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 6
US-10-074-225A-1
; Sequence 1, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; FILE OF INVENTION: ANTI-TUMOR AGENT

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2005, 09:20:45 ; Search time 44 Seconds
(without alignments)
61.077 Million cell updates/sec

Title: US-10-500-018A-27

Perfect score: 215

Sequence: 1 EPHHHFGQRMHGYDADVRLYRRHHGSGSPSRHRH 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.5	46.7	344	1	US-08-657-192-3
2	100.5	46.7	344	3	US-08-523-373-5
3	100.5	46.7	392	3	US-08-523-373-6
4	100.5	46.7	532	1	US-08-657-192-9
5	100.5	46.7	532	3	US-08-523-373-7
6	100.5	46.7	537	1	US-08-657-192-15
7	72	33.5	465	4	US-09-252-991A-24848
8	70	32.6	15	2	US-08-811-028-52
9	69	32.1	556	4	US-09-252-991A-21219
10	68.5	31.9	736	4	US-09-252-991A-32256
11	67.5	31.4	357	4	US-09-252-991A-30141
12	67.5	31.4	890	4	US-09-252-991A-19564
13	66	30.7	525	4	US-09-976-594-64
14	66	30.7	525	4	US-09-919-039-62
15	66	30.7	1014	4	US-09-252-991A-29868
16	64.5	30.0	398	4	US-09-252-991A-19488
17	63	29.3	351	4	US-09-252-991A-18551
18	63	29.3	463	4	US-09-252-991A-28820
19	62.5	29.1	189	4	US-09-252-991A-17056
20	62.5	29.1	580	4	US-09-252-991A-22036
21	62	28.8	425	4	US-09-252-991A-25840
22	62	28.8	429	4	US-09-252-991A-33116
23	61.5	28.6	252	4	US-09-252-991A-18141
24	61.5	28.6	437	4	US-09-489-039A-13489
25	61.5	28.6	724	4	US-09-252-991A-31715
26	61	28.4	834	4	US-09-252-991A-18401
27	60.5	28.1	392	4	US-09-252-991A-21863

28	60.5	28.1	591	4	US-09-252-991A-31286	Sequence 31286, A
29	60.5	28.1	952	4	US-09-252-991A-32183	Sequence 32183, A
30	60	27.9	89	3	US-09-507-323B-12	Sequence 12, Appl
31	60	27.9	167	3	US-09-507-323B-13	Sequence 13, Appl
32	60	27.9	437	4	US-09-252-991A-28176	Sequence 28176, A
33	60	27.9	1402	4	US-09-248-796A-14503	Sequence 14503, A
34	59.5	27.7	170	4	US-09-252-991A-27727	Sequence 27727, A
35	59.5	27.7	295	4	US-09-489-039A-9496	Sequence 9496, Ap
36	59.5	27.7	359	4	US-09-252-991A-17283	Sequence 17283, A
37	59.5	27.7	579	4	US-09-540-236-2373	Sequence 2373, Ap
38	59	27.4	142	4	US-09-252-991A-28784	Sequence 28784, A
39	59	27.4	311	4	US-09-252-991A-18075	Sequence 18075, A
40	59	27.4	374	4	US-09-252-991A-33040	Sequence 33040, A
41	59	27.4	375	4	US-09-489-039A-9924	Sequence 9924, Ap
42	59	27.4	689	4	US-09-252-991A-19258	Sequence 19258, A
43	58.5	27.2	310	4	US-09-252-991A-27948	Sequence 27948, A
44	58.5	27.2	473	4	US-09-252-991A-24399	Sequence 24399, A
45	58.5	27.2	525	4	US-09-252-991A-31946	Sequence 31946, A

ALIGNMENTS

RESULT 1

US-08-657-192-3
; Sequence 3, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: YABUTA, Masayuki
; APPLICANT: OHSUYE, Kazuhiro
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DONNE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Neuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-657-192-3

Query Match 46.7%; Score 100.5; DB 1; Length 344;
Best Local Similarity 66.7%; Pred. No. 3e-05; 5; Indels 3; Gaps 1;
Matches 20; Conservative 2; Mismatches 3; Gaps 1;

QY 7 PGGRMHGYDADVRLYRRHH---GSGSPSR 33

```

Db      88 PSNQMHHGYDAELRLYRRHHRWGSGSPLR 117

RESULT 2
US-08-523-373-5
; Sequence 5, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohshuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuch, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-5

Query Match 46.7%; Score 100.5; DB 3; Length
Best Local Similarity 66.7%; Pred. No.3e-05; 5; Indel
Matches 20; Conservative

QY      7 PGCGRMHGYDADRLYRRHH---CGSGSPSR 33
DB      88 PSNQMHHGYDAELRLYRRHHRWGSGSPLR 117

RESULT 3
US-08-523-373-6
; Sequence 6, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohshuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria

```